

155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973

[illegible]

distribution. We also found information can be found
 at the I.M.A.G.E. Center located at: <http://image.umd.gov>
 IPK plate: 42 Row: 8 Column: 7
 one has the following problem: no 5' EST match.
 Location/Qualifiers:

```

1: 3821
2: filename="Hong_sai.jpg"
3: xlabel="xaved_96"
4: ylabel="yaved_96"
5: csize=IMAGE(4,3,255)
6: csize_type="uint8_t", csize_order="C"
7: img_loss
8: img_lib="Net_SMB"
9: img_lib_order="H"
10: img_lib_order="H"
11: img_lib_order="H"
12: img_lib_order="H"
13: img_lib_order="H"

```

49.2%; Score 4.4; OR 11; Length 2821;
 affinity 80.9%; Eval No. 4,000,142;
 sensitive 0; Mismatches 225; indels 13; Gaps 3;
 5' GAGTAACTGCTCCGACAGGAGCTTTGGGCTGCTGCTGATTTTAAATT 25
 |||||
 3' ATTTTGGGCGACAGGCTTGGTGTGACGAGGCTGTAATTTGCGAATT 62
 |||||
 5' GATGACATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 135
 |||||
 3' AGTATGTAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 116
 |||||
 5' TTGTTGGGAGACCAATGCTTTTAAATGCTGAGCTCCAGAAATATGCTACC 195
 |||||
 3' TTGTTGGGAGACCAATGCTTTTAAATGCTGAGCTCCAGAAATATGCTACC 176
 |||||

[illegible][illegible]

RESULT 3	AL537523	1608 bp	cdna	1186aa	EST 13 FEB 2001
LOCUS	AL537523	LT1 FL013	Homo sapiens cDNA clone (c0406246)E18.5		
DEFINITION	AL537523	prime, mRNA sequence.			
ACCESSION	AL537523				
VERSION	AL537523.1	GI:12931016			
KEYWORDS	EST.				
SOURCE	human				
ORGANISM	Homo sapiens				
REFERENCE	1. (Bases 1 to 1005)				
AUTHORS	G.W.B., Gruber,C., Jesse,J. and Polyes,D.				
TITLE	Full length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Genoscope - Centre National de Sequencage				




```

DB 334 GAAGTTCAGAGAACTATCTATGAAAGAGAAAGAGGCTATGCTAGGCTGTCTCTGATGGGTTCC 393
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
OY 415 TGGGAGGAGGTAAGAGCCCTGGCTGAAACAGACCTGTGATGAAGTTCTA 461
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 394 TGGGAGGAGGTAAGAGCCCTGGCTGAAACAGACCTGTGATGAAGTTCTA 440
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

Search completed: April 7, 2003, 00:34:12
Job time : 99 secs

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

... of ...

```

Db 1561 CCGGACCTGCGGAGTCTGGAATGGAACACCGCTGCTGATGAGATCATGAGAGATGAGGCTG 1648
Cy 1635 TGGACATTAATCAATAAAGAACAGAGATTAATGAGAAATGAGAAATGAGAAATGAGAAAT 1684
Gb 1621 TGGCAATTAATCAATAAAGAACAGAGATTAATGAGAAATGAGAAATGAGAAATGAGAAAT 1670

RESULT 4
ARN99656
ID ABN99656 standard; DNA; 1648 BP.
Ac ARN99656;
Xt 16-AUG-2002 (first entry)
Human clusterin gene sequence 1.
Xw Human; antisense inhibition; antisense oligonucleotide; clusterin;
Xw hypercholesterolemia; small-molecular inhibitor; de;
Xw hyperproliferative disorder; hyperlipidemic disorder.
Xx Homo sapiens.
Xx WO200222635 A1.
Xn 21-MAR-2002.
Xx 10-SEP-2001; 2001WO 0029235.
Xr 11-SEP-2000; 2000US-0659791.
Xs (ISIS-) ISIS PHARM INC.
Xa Monia BP, Fraier SM;
Xx WPI; 2002-404805/43.
Xr P-PSDB; AA015019.
Xx Novel antisense compound targeted to human and murine clusterin
Xr clusterin; useful for treating animal having disease associated with
Xr clusterin such as hyperlipidemic disorder; cardiovascular disorder
Xx Example 1); Page 99-102; 1999, English.
Xs The invention describes antisense oligonucleotides that are capable of
Xc inhibiting expression of the human clusterin gene. The antisense
Xc oligonucleotides of the invention are useful for inhibiting the
Xc expression of clusterin in cells. The antisense oligonucleotides are also
Xc useful for treating an animal with a disease or condition associated with
Xc clusterin (e.g., hypercholesterolemia, cardiovascular disorders;
Xc hyperproliferative disorders; and hyperlipidemic disorders). The present
Xc DNA sequence represents a human clusterin gene sequence.
Xx sequence 1648 BP; 408 A; 448 G; 440 C; 412 T; 0 other;

Query Match 60.0%; Score 1010.8; BB 24; Length 1648;
Best Local Similarity 79.4%; E-Val 9.1e-322;
Matches 1295; Conservative 7; Gaps 7;

Cy 39 GGCACGCTTAACGAGAGGCTGCTGGATTCGASATTCGAGAGGAGGAGGACATGAGAC 98
Db 10 GGTGACGCTTAACGAGAGGAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 63
Cy 69 TGTGATGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 158
Db 64 TGTGATGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 123
Cy 159 GATGCTGCTGAGACATGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 218
Db 124 GACGCTGCTGAGACATGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 183
Cy 219 GGAATTAATTAATGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 278

```

```

Db 1281 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243
Cy 275 CCGAGAGAGGAGATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
Gb 234 CCGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
Cy 329 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
Db 304 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
Cy 359 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
Db 364 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
Cy 459 CTATGCGCGAGCTGCGAGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
Db 424 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
Cy 519 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
Db 484 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
Cy 579 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Db 544 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Cy 639 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
Db 604 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
Cy 699 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 758
Db 664 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 758
Cy 759 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 818
Db 724 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 818
Cy 819 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
Db 784 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
Cy 879 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
Db 844 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
Cy 939 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
Db 904 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
Cy 999 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1058
Db 964 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1058
Cy 1059 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
Db 1024 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
Cy 1119 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1178
Db 1084 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1178
Cy 1179 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1238
Db 1144 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1238
Cy 1239 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
Db 1204 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
Cy 1299 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1358
Db 1264 CCGAATTTGAAATGCTGCTGAGAGTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1358

```



```

* 7377 7476: gap of unknown length
* 7477 8725: contig of 1249 bp in length
* 8726 8823: gap of unknown length
* 8823 10186: contig of 1663 bp in length
* 10186 10788: gap of unknown length
* 10788 11454: contig of 1166 bp in length
* 11454 11554: gap of unknown length
* 11554 13173: contig of 1519 bp in length
* 13173 13273: gap of unknown length
* 13273 14432: contig of 1159 bp in length
* 14432 14532: gap of unknown length
* 14532 15798: contig of 1266 bp in length
* 15798 15898: gap of unknown length
* 15898 16373: contig of 1031 bp in length
* 16373 16779: gap of unknown length
* 16779 18647: contig of 1569 bp in length
* 18647 18747: gap of unknown length
* 18747 19044: contig of 1097 bp in length
* 19044 19344: gap of unknown length
* 19344 21319: contig of 1174 bp in length
* 21319 21418: gap of unknown length
* 21418 22571: contig of 1492 bp in length
* 22571 23171: gap of unknown length
* 23171 24188: contig of 1187 bp in length
* 24188 24288: gap of unknown length
* 24288 25591: contig of 1112 bp in length
* 25591 25690: gap of unknown length
* 25690 27143: contig of 1453 bp in length
* 27143 27243: gap of unknown length
* 27243 28343: contig of 1106 bp in length
* 28343 29415: gap of unknown length
* 29415 29662: contig of 1213 bp in length
* 29662 30720: gap of unknown length
* 30720 31161: contig of 1399 bp in length
* 31161 32370: contig of 1109 bp in length
* 32370 32470: gap of unknown length
* 32470 34107: contig of 1637 bp in length
* 34107 34277: gap of unknown length
* 34277 35394: contig of 1187 bp in length
* 35394 35494: gap of unknown length
* 35494 35896: contig of 1402 bp in length
* 35896 36476: gap of unknown length
* 36476 38359: contig of 1363 bp in length
* 38359 38459: gap of unknown length
* 38459 39572: contig of 1233 bp in length
* 39572 40093: gap of unknown length
* 40093 42033: gap of unknown length
* 42033 43161: contig of 2068 bp in length
* 43161 43761: gap of unknown length
* 43761 44459: contig of 1199 bp in length
* 44459 45927: contig of 1268 bp in length
* 45927 45927: gap of unknown length
* 45927 47186: contig of 1259 bp in length
* 47186 48286: gap of unknown length
* 48286 49071: contig of 1695 bp in length
* 49071 50341: gap of unknown length
* 50341 50441: gap of unknown length
* 50441 51855: contig of 1414 bp in length
* 51855 51956: gap of unknown length
* 51956 53089: contig of 1134 bp in length
* 53089 53189: gap of unknown length
* 53189 54954: contig of 1765 bp in length
* 54954 55491: gap of unknown length
* 55491 56491: contig of 1437 bp in length
* 56491 56591: gap of unknown length
* 56591 58292: contig of 1701 bp in length
* 58292 58392: gap of unknown length
* 58392 59309: contig of 1517 bp in length
* 59309 60009: gap of unknown length

```

```

* 60010 61437: contig of 1428 bp in length
* 61437 61571: gap of unknown length
* 61571 64033: contig of 2496 bp in length
* 64033 64133: gap of unknown length
* 64133 65784: contig of 1572 bp in length
* 65784 65804: gap of unknown length
* 65804 67677: contig of 1874 bp in length
* 67677 67777: gap of unknown length
* 67777 69446: contig of 1669 bp in length
* 69446 69546: gap of unknown length
* 69546 70900: contig of 1354 bp in length
* 70900 71000: gap of unknown length
* 71000 73231: contig of 2231 bp in length
* 73231 73331: gap of unknown length
* 73331 74362: contig of 1031 bp in length
* 74362 74462: gap of unknown length
* 74462 76534: contig of 2072 bp in length
* 76534 76731: gap of unknown length
* 76731 78509: contig of 1875 bp in length
* 78509 78609: gap of unknown length
* 78609 80611: contig of 2002 bp in length
* 80611 82711: gap of unknown length
* 82711 84492: contig of 1698 bp in length
* 84492 85441: contig of 1041 bp in length
* 85441 85541: gap of unknown length
* 85541 85641: gap of unknown length
* 85641 85741: gap of unknown length
* 85741 85841: gap of unknown length
* 85841 85941: gap of unknown length
* 85941 86041: gap of unknown length
* 86041 86141: gap of unknown length
* 86141 86241: gap of unknown length
* 86241 86341: gap of unknown length
* 86341 86441: gap of unknown length
* 86441 86541: gap of unknown length
* 86541 86641: gap of unknown length
* 86641 86741: gap of unknown length
* 86741 86841: gap of unknown length
* 86841 86941: gap of unknown length
* 86941 87041: gap of unknown length
* 87041 87141: gap of unknown length
* 87141 87241: gap of unknown length
* 87241 87341: gap of unknown length
* 87341 87441: gap of unknown length
* 87441 87541: gap of unknown length
* 87541 87641: gap of unknown length
* 87641 87741: gap of unknown length
* 87741 87841: gap of unknown length
* 87841 87941: gap of unknown length
* 87941 88041: gap of unknown length
* 88041 88141: gap of unknown length
* 88141 88241: gap of unknown length
* 88241 88341: gap of unknown length
* 88341 88441: gap of unknown length
* 88441 88541: gap of unknown length
* 88541 88641: gap of unknown length
* 88641 88741: gap of unknown length
* 88741 88841: gap of unknown length
* 88841 88941: gap of unknown length
* 88941 89041: gap of unknown length
* 89041 89141: gap of unknown length
* 89141 89241: gap of unknown length
* 89241 89341: gap of unknown length
* 89341 89441: gap of unknown length
* 89441 89541: gap of unknown length
* 89541 89641: gap of unknown length
* 89641 89741: gap of unknown length
* 89741 89841: gap of unknown length
* 89841 89941: gap of unknown length
* 89941 90041: gap of unknown length
* 90041 90141: gap of unknown length
* 90141 90241: gap of unknown length
* 90241 90341: gap of unknown length
* 90341 90441: gap of unknown length
* 90441 90541: gap of unknown length
* 90541 90641: gap of unknown length
* 90641 90741: gap of unknown length
* 90741 90841: gap of unknown length
* 90841 90941: gap of unknown length
* 90941 91041: gap of unknown length
* 91041 91141: gap of unknown length
* 91141 91241: gap of unknown length
* 91241 91341: gap of unknown length
* 91341 91441: gap of unknown length
* 91441 91541: gap of unknown length
* 91541 91641: gap of unknown length
* 91641 91741: gap of unknown length
* 91741 91841: gap of unknown length
* 91841 91941: gap of unknown length
* 91941 92041: gap of unknown length
* 92041 92141: gap of unknown length
* 92141 92241: gap of unknown length
* 92241 92341: gap of unknown length
* 92341 92441: gap of unknown length
* 92441 92541: gap of unknown length
* 92541 92641: gap of unknown length
* 92641 92741: gap of unknown length
* 92741 92841: gap of unknown length
* 92841 92941: gap of unknown length
* 92941 93041: gap of unknown length
* 93041 93141: gap of unknown length
* 93141 93241: gap of unknown length
* 93241 93341: gap of unknown length
* 93341 93441: gap of unknown length
* 93441 93541: gap of unknown length
* 93541 93641: gap of unknown length
* 93641 93741: gap of unknown length
* 93741 93841: gap of unknown length
* 93841 93941: gap of unknown length
* 93941 94041: gap of unknown length
* 94041 94141: gap of unknown length
* 94141 94241: gap of unknown length
* 94241 94341: gap of unknown length
* 94341 94441: gap of unknown length
* 94441 94541: gap of unknown length
* 94541 94641: gap of unknown length
* 94641 94741: gap of unknown length
* 94741 94841: gap of unknown length
* 94841 94941: gap of unknown length
* 94941 95041: gap of unknown length
* 95041 95141: gap of unknown length
* 95141 95241: gap of unknown length
* 95241 95341: gap of unknown length
* 95341 95441: gap of unknown length
* 95441 95541: gap of unknown length
* 95541 95641: gap of unknown length
* 95641 95741: gap of unknown length
* 95741 95841: gap of unknown length
* 95841 95941: gap of unknown length
* 95941 96041: gap of unknown length
* 96041 96141: gap of unknown length
* 96141 96241: gap of unknown length
* 96241 96341: gap of unknown length
* 96341 96441: gap of unknown length
* 96441 96541: gap of unknown length
* 96541 96641: gap of unknown length
* 96641 96741: gap of unknown length
* 96741 96841: gap of unknown length
* 96841 96941: gap of unknown length
* 96941 97041: gap of unknown length
* 97041 97141: gap of unknown length
* 97141 97241: gap of unknown length
* 97241 97341: gap of unknown length
* 97341 97441: gap of unknown length
* 97441 97541: gap of unknown length
* 97541 97641: gap of unknown length
* 97641 97741: gap of unknown length
* 97741 97841: gap of unknown length
* 97841 97941: gap of unknown length
* 97941 98041: gap of unknown length
* 98041 98141: gap of unknown length
* 98141 98241: gap of unknown length
* 98241 98341: gap of unknown length
* 98341 98441: gap of unknown length
* 98441 98541: gap of unknown length
* 98541 98641: gap of unknown length
* 98641 98741: gap of unknown length
* 98741 98841: gap of unknown length
* 98841 98941: gap of unknown length
* 98941 99041: gap of unknown length
* 99041 99141: gap of unknown length
* 99141 99241: gap of unknown length
* 99241 99341: gap of unknown length
* 99341 99441: gap of unknown length
* 99441 99541: gap of unknown length
* 99541 99641: gap of unknown length
* 99641 99741: gap of unknown length
* 99741 99841: gap of unknown length
* 99841 99941: gap of unknown length

```


[illegible]


```
17 108 5.7 916 5 Q9RUC6
18 108 5.7 798 5 Q1763
19 107 5.6 432 10 Q81G2
20 107 5.6 7231 5 Q9MT8
21 106 5.5 496 4 Q8755
22 105 5.5 1738 5 Q8325
23 104 5.4 1833 2 Q9L89
24 104 5.4 1623 5 Q9307
25 103.5 5.4 363 11 Q9EQ9
26 103.5 5.4 625 6 Q8717
27 103.5 5.4 2867 5 Q8X23
28 103.5 5.4 3560 4 Q8X28
29 103.5 5.4 5351 4 Q8X29
30 103 5.4 567 10 Q91X2
31 103 5.4 774 16 Q32215
32 103 5.4 1780 4 Q9QF5
33 103 5.4 1820 4 Q9Q89
34 102 5.3 1266 4 Q8T21
35 102 5.3 990 5 Q8VES
36 100.5 5.3 439 10 Q9FJ7
37 100.5 5.3 1173 4 Q9U86
38 100.5 5.3 1312 4 Q92878
39 100.5 5.3 1318 4 Q43254
40 100.5 5.3 1728 10 Q9LH2
41 100 5.2 651 12 Q9DVI2
42 100 5.2 1217 4 Q60454
43 100 5.2 1217 11 Q9QUS1
44 99.5 5.2 548 2 Q9A74
45 99.5 5.2 1388 4 Q9NS87
```

ALIGNMENTS

RESULT 1

```
Q29482 ID Q29482 PRELIMINARY: PRT: 449 AA.
AC Q29482;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Clusterin precursor.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
OF Mammalia; Perissodactyla; Equidae; Equus.
OX NCBI TaxID 9796;
RN [1]
RE SEQUENCE FROM N.A.
RC SPAIN-OLIVER; TISSUE-TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Craibe G.J., Foster D.J.
RT "Nucleotide sequence of the complementary DNA encoding Equine
Clusterin."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO ELISA, MEMBRANES,
ANT-HYPERHEMATIC PROTEINS. IT HAS BEEN ASSOCIATED WITH INFLAMED
CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTI-PARALLEL DISULFIDE-LINKED HEIMER OLIGOMER (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL: L46797; AAA80313.1;
DR InterPro: IP000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; CLA; 1.
DR SMART: SM00030; CLE; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
CC SEQUENCE 449 AA, 5195 MW, 103450 Da, 103450 Da.
Query Match 76.9% Score 1490.5; DP 4; Length 449;
```

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

bioactive peptides (11411, 4978
1993-2003) compiled by:

CC AND HYDROPHILIC PROTEINS. IT HAS BEEN ASSOCIATED WITH INCREASED
 CC CELL DEATH (BY SIMILARITY).
 CC : SCURVIT, ANGIOTENSIN II RECEPTOR-LIKE RECEPTOR (BY
 CC SIMILARITY).
 CC : SIMILARITY. BELONGS TO THE CLUSTERIN FAMILY.
 DR EMBL; AF14657; AAG31462.1; -
 DE InterPro; IPR00093; Clusterin.
 DR Pfam; PF01093; Clusterin; 1.
 DR SMART; SM00030; CLB; 1.
 DR PROSITE; PS00493; CLUSTERIN_2; 1.
 KW Glycoprotein.
 FT NON TER.
 ST SEQUENCE 126 AA; 54261 MW; 141144.242 kDa; 45.94.

Query Match 16.81; Score 320; DB 47; Length 466;
 Best Local Similarity 22.61; Ids: 82; 95 16;
 Matches 96; Conservative 90; Mismatches 125; Indels 59; Gaps 12;

CC 18: EMBL; AF14657; AAG31462.1; -
 CC 1: RASGIDILFOURHFEQHHFSP-----
 CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 18: EMBL; AF14657; AAG31462.1; -
 CC 1: RASGIDILFOURHFEQHHFSP-----
 CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

CC 28: -----
 CC 29: EMBL; AF14657; AAG31462.1; -
 CC 72: EMBL; AF14657; AAG31462.1; -

QV 61 QTR-----ERKSLTNGEAFPRFELATFSEMFDEACVNTVMWALKEEFP 114
D 245 ERMEMVWTEGROVIAH RLLPLAEIVAS--LLEPFUELVGHEMNEH 47
QV 115 CL-----KOTCMFYAVVRSSTGLVGHVHFHFNUNTFYWNISDFD 149
D 318 SMKESNMVEITAFOPFTFSEVEFEGENLIERMEMEYVSSD KRGID 412
QV 150 -----SLLENCR--GCHALIMQSFQFASIMFEL 189
D 373 CLERKEVELEFICFHFVLEINFEALITHAVATLQNTLCHINHL 421
RESULT 15
QRMXHO
AC QRMXHO IALLIMINAFY 151, 4895 AA.
DT 01-MAR-2002 (Temblor): 20, Created
DT 01-MAR-2002 (Temblor): 20, Last sequence update
DT 01-JUN-2002 (Temblor): 21, Last annotation update
DE NIANCE.
GN NUA.
OS Homo sapiens (Human).
CC Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi;
CC Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
OX NCBI_TaxID=9606;
RN [1]
RF SEQUENCE FROM N.A.
RA Zhen Y.-Y., Tibotol T., Nohari A.A., Kothachan R.J.
RT NUAACE, a giant protein connecting nucleus and cytoskeleton;
RL Submitted (02-2001) to the EMBL/GenBank/CCDS databases.
ER EXEL, AF435011; AAL3354811;
DR InterPro; IPR001589; Actbind actin.
DR InterPro; IPR000515; PPD transp.
DR InterPro; IPR001715; Caponin-like.
DR InterPro; IPR005144; Ektel receptor.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00810; ER_lumen_recept; 1.
DR Pfam; PF00435; spectrin; 14.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 20.
DR PROSITE; PS00014; ACTININ_1; UNKN:WN 1.
DR PROSITE; PS00442; EFD_TRANSF_INN_MEMER; UNKN:WN 1.
DR PROSITE; PSS0031; CH; 2.
SQ SEQUENCE 4895 AA; 74665 MW; 1176555-444444 P 147;
Query Match 5.7%; Score 109; DP 4; Identity 100%;
Best Local Similarity 21.6%; Pred. No. 62;
Marches 92, Conservation 60, Mismatch 111, InDel 89, Gaps 22;
QV 75 VQTF--LTPVTEFVVI--WEIPNHFVVFVRIHED--TNRD 65
D 1410 LSEIHVYVAEETIEEMLLEAFIPKNNEL--LEKIVQVQVNS--LKEFIVIAVHP 1466
QV 55 PPSL--LTLNEAPPRKETAAL--KSEWELASD--NVTWMAWKEEF 113
D 1467 KKSIRLEKLVDEYEREPHLCGMANSLHFFGSPETNCC--INTV--LWNTFYALV 1624
QV 114 -PCIKQ-----TCKFYAVVCRSTGLVGHVHFHFNUNTFYWNISDFD 149
D 1624 IECLEPQSVLEIKVLAHPSHILILL--PEESVNS--GAYM-----GEMLR 1672
QV 166 QTHALIMQSFQFASIMFEL--GCHALIMQSFQFASIMFEL 189
D 1672 KIALIEVPEEFHELEWKEING--VKNLQ--FYLEMKRTHGFF--LRIANI 1622
QV 359 IMAFHSGFCEINLMPGPIELMHQA--GNGVNHSLH--PMH--LKN--GAV--KEIRH 287
D 1682 LVS--SWLIERKILVLEHGFALAE--ELNATVIL--L--AL--EMELH 1672
QV 388 NSTCLDQVQCVTELEIVS----SNNEATV--FLEKNSLIEP--FTFVAV 449
D 1692 LVS--SWLIERKILVLEHGFALAE--ELNATVIL--L--AL--EMELH 1672

QV 61 QTR-----ERKSLTNGEAFPRFELATFSEMFDEACVNTVMWALKEEFP 114
D 245 ERMEMVWTEGROVIAH RLLPLAEIVAS--LLEPFUELVGHEMNEH 47
QV 115 CL-----KOTCMFYAVVRSSTGLVGHVHFHFNUNTFYWNISDFD 149
D 318 SMKESNMVEITAFOPFTFSEVEFEGENLIERMEMEYVSSD KRGID 412
QV 150 -----SLLENCR--GCHALIMQSFQFASIMFEL 189
D 373 CLERKEVELEFICFHFVLEINFEALITHAVATLQNTLCHINHL 421
RESULT 15
QRMXHO
AC QRMXHO IALLIMINAFY 151, 4895 AA.
DT 01-MAR-2002 (Temblor): 20, Created
DT 01-MAR-2002 (Temblor): 20, Last sequence update
DT 01-JUN-2002 (Temblor): 21, Last annotation update
DE NIANCE.
GN NUA.
OS Homo sapiens (Human).
CC Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi;
CC Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
OX NCBI_TaxID=9606;
RN [1]
RF SEQUENCE FROM N.A.
RA Zhen Y.-Y., Tibotol T., Nohari A.A., Kothachan R.J.
RT NUAACE, a giant protein connecting nucleus and cytoskeleton;
RL Submitted (02-2001) to the EMBL/GenBank/CCDS databases.
ER EXEL, AF435011; AAL3354811;
DR InterPro; IPR001589; Actbind actin.
DR InterPro; IPR000515; PPD transp.
DR InterPro; IPR001715; Caponin-like.
DR InterPro; IPR005144; Ektel receptor.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00810; ER_lumen_recept; 1.
DR Pfam; PF00435; spectrin; 14.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 20.
DR PROSITE; PS00014; ACTININ_1; UNKN:WN 1.
DR PROSITE; PS00442; EFD_TRANSF_INN_MEMER; UNKN:WN 1.
DR PROSITE; PSS0031; CH; 2.
SQ SEQUENCE 4895 AA; 74665 MW; 1176555-444444 P 147;
Query Match 5.7%; Score 109; DP 4; Identity 100%;
Best Local Similarity 21.6%; Pred. No. 62;
Marches 92, Conservation 60, Mismatch 111, InDel 89, Gaps 22;
QV 75 VQTF--LTPVTEFVVI--WEIPNHFVVFVRIHED--TNRD 65
D 1410 LSEIHVYVAEETIEEMLLEAFIPKNNEL--LEKIVQVQVNS--LKEFIVIAVHP 1466
QV 55 PPSL--LTLNEAPPRKETAAL--KSEWELASD--NVTWMAWKEEF 113
D 1467 KKSIRLEKLVDEYEREPHLCGMANSLHFFGSPETNCC--INTV--LWNTFYALV 1624
QV 114 -PCIKQ-----TCKFYAVVCRSTGLVGHVHFHFNUNTFYWNISDFD 149
D 1624 IECLEPQSVLEIKVLAHPSHILILL--PEESVNS--GAYM-----GEMLR 1672
QV 166 QTHALIMQSFQFASIMFEL--GCHALIMQSFQFASIMFEL 189
D 1672 KIALIEVPEEFHELEWKEING--VKNLQ--FYLEMKRTHGFF--LRIANI 1622
QV 359 IMAFHSGFCEINLMPGPIELMHQA--GNGVNHSLH--PMH--LKN--GAV--KEIRH 287
D 1682 LVS--SWLIERKILVLEHGFALAE--ELNATVIL--L--AL--EMELH 1672
QV 388 NSTCLDQVQCVTELEIVS----SNNEATV--FLEKNSLIEP--FTFVAV 449
D 1692 LVS--SWLIERKILVLEHGFALAE--ELNATVIL--L--AL--EMELH 1672

cc between the Swiss Institute of Bioinformatics and the EMBL outstation
 cc the European Bioinformatics Institute. There are no restrictions on its
 cc use by non-profit institutions as long as its content is in no way
 cc modified and this statement is not removed. Usage by and for commercial
 cc entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 cc or send an email to license@isb-sib.ch).

DR EMBL; AL133357; CAB62414.1; ..
 DR EMBL; AF010473; AAB65416.1; ALT_INIT.
 DR EMBL; AB028012; BAAB7316.1; ..
 KW Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1327 COILED COIL (POTENTIAL).
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1727 AA; 197856 MW; F820RPM69G132644 CRC64;

Query Match 5.4%; Score 102.5; DB 1; Length 1727;
 Best Local Similarity 19.5%; Pred. No. 14;
 Matches 83, Conservation 69, Mismatches 172, Indels 103, Gaps 18;

OY 10 NQVDELMWCTELC EMTEEDGVINFEI KWALKG 51
 DB 1203 NPLPRLGELGELGELKLVAVVLENSREYVHHHLLHLSSTGTCGACPLPHELV 1328
 OY 50 VPCITPTIETITPPPPGILNIEAPPPPEY ALNTPPEVPLPAQVYVETMM 156
 DB 1329 E ENLIEELNCELEHPELEETVKECELEENHAGNVEILEL NPLA 1079
 OY 107 ALW FEGHGLKLTOMKPVAVVAVCSSTGLVAGHVEELN----- 144
 DB 1380 AAMPTVYEVVTPSSEVYVHCTEQSEVTELEAEVAC ----HNLNLPDSIPT 1433
 OY 145 -GNSPPVFWIN-GRPLNLEMLPQVCHALLWJUSFDEASSIMELFQGFPTFEA 199
 DB 1434 ATTLEHGVLEENSHLEELLTFLCEHNGVHLLTLEELVAVNCHLESTET 1493
 OY 233 LIDHHTPDGDSIEIFF FHHHFAHNMHTTGHLEN FHEMTFFETENHQA 253
 DB 1494 EIDNF FHWPPPPVAVPTVETETETAPTTPPTSTVQVNLTVVASEETGCS 1549
 OY 254 LQAMVNLHLLHSMEFELMLQVAVPELSH N SLDLE MPELEPER 303
 DB 1550 EEARASVH AHTHELENLETEENMLSVTFQVPHESISESTSTAGTTPVCHWEER 1608
 OY 304 EILSVNCEENNAQVLEF ELNLSGLIA EYTELVEEAAVLPDEYQHV 383
 DB 1609 ELLPVLLEVALQHLKLENNERELEENHLEENLEHLEAVFAELEETETETIAI 1667
 OY 354 LPABAA 359
 DB 1668 LSLEAS 1673

Search completed: April 2, 2003, 13:53:17
 Job time : 18 secs

Query Match 38.38; Score 741.54; LR 2; Length 448;
Best Local Similarity 42.68; Pred. No. 146-45;
Matched 147; Conserving 12; Mismatched 105; Indels 21; Gaps 7;
Query 4 LLLVGLLTWNGRVIGDQVSDTEQENSTGSGVKNKKEINAKLGGVQIKTLKTN 60
D1 5 LLLVGLLTWNGRVIGDQVSDTEQENSTGSGVKNKKEINAKLGGVQIKTLKTN 60
QY 64 EERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILN 123
D1 64 EERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILN 123
QY 61 PERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTE 120
D1 61 PERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTE 120
QY 124 YAPVCRSTGLVCHQVEEPFVCSGIVGVNINDFIRGLLENFVCTHALVQVDSIFAS 193
D1 124 YAPVCRSTGLVCHQVEEPFVCSGIVGVNINDFIRGLLENFVCTHALVQVDSIFAS 193
QY 121 YSPVCHSGSLVAGQLKEELNLSAPFELWUBERIGALIPPOFSEPPFHLSEPPHME 180
D1 121 YSPVCHSGSLVAGQLKEELNLSAPFELWUBERIGALIPPOFSEPPFHLSEPPHME 180
QY 184 SIMDELQDPPFFREAGDP---FHPSPSSPFRPPFNI---KHPFAPNIMPF 212
D1 184 SIMDELQDPPFFREAGDP---FHPSPSSPFRPPFNI---KHPFAPNIMPF 212
QY 181 NVGDIFFQ---STQVGPAPFFFTTPPESTF REAFVIFVGVVHVPFELSECHIFF 216
D1 181 NVGDIFFQ---STQVGPAPFFFTTPPESTF REAFVIFVGVVHVPFELSECHIFF 216
QY 233 GYGLNLNLEMEHFFPMI-HUAGGAMVNLHPLHHEIMFFELHAGQSAV KEIHNSTG 291
D1 233 GYGLNLNLEMEHFFPMI-HUAGGAMVNLHPLHHEIMFFELHAGQSAV KEIHNSTG 291
QY 237 QHPHGHRLPFLQFEMTQHMLOGGHGANEHPLGGGATF-SNFATTPAV PEIHNNA 295
D1 237 QHPHGHRLPFLQFEMTQHMLOGGHGANEHPLGGGATF-SNFATTPAV PEIHNNA 295
QY 295 TTPWYTSFTFTTITVWTSQDAVAVLGLNLNLEMEHFFELHAGQSAV KEIHNSTG 326
D1 295 TTPWYTSFTFTTITVWTSQDAVAVLGLNLNLEMEHFFELHAGQSAV KEIHNSTG 326
QY 296 CLEMEDCEKCPREILAVDCQCTFVSGLPFEPEDALPLAEPTF 340
D1 296 CLEMEDCEKCPREILAVDCQCTFVSGLPFEPEDALPLAEPTF 340

RESULT 9
150131
clusterin - quat
Accession: 150131
C:Date: 13-Sep-1996 #sequence revision 13 Sep 1996 #next change 13 Aug 1999
R:Michel, D.; Chacelain, S.; Herault, Y.; Brun, G.
Eur. J. Biochem. 229, 215-223, 1995
Article: the expression of the avian clusterin gene has been cloned, two alternative promoters identified, and the structure of the gene determined.
A:Accession: 150131
A:Reference: 150131, NCBI:55262678; PMID:77441
A:Status: preliminary; translated from cDNA
A:Molecule type: DNA
A:Residues: 1451; <MC>
A:Cross-references: EMBL:X80740; NID:946299; PIR:TAAG033.1; FID:42940
A:Gene: T44
A:Superfamily: clusterin

Query Match 38.38; Score 741.54; LR 2; Length 448;
Best Local Similarity 42.68; Pred. No. 146-45;
Matched 147; Conserving 12; Mismatched 105; Indels 21; Gaps 7;
Query 4 LLLVGLLTWNGRVIGDQVSDTEQENSTGSGVKNKKEINAKLGGVQIKTLKTN 60
D1 5 LLLVGLLTWNGRVIGDQVSDTEQENSTGSGVKNKKEINAKLGGVQIKTLKTN 60
QY 64 EERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILN 123
D1 64 EERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILNLEERPSILN 123
QY 61 PERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTE 120
D1 61 PERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTEETPERHAMLTE 120
QY 124 YAPVCRSTGLVCHQVEEPFVCSGIVGVNINDFIRGLLENFVCTHALVQVDSIFAS 193
D1 124 YAPVCRSTGLVCHQVEEPFVCSGIVGVNINDFIRGLLENFVCTHALVQVDSIFAS 193
QY 121 YSPVCHSGSLVAGQLKEELNLSAPFELWUBERIGALIPPOFSEPPFHLSEPPHME 180
D1 121 YSPVCHSGSLVAGQLKEELNLSAPFELWUBERIGALIPPOFSEPPFHLSEPPHME 180
QY 184 SIMDELQDPPFFREAGDP---FHPSPSSPFRPPFNI---KHPFAPNIMPF 212
D1 184 SIMDELQDPPFFREAGDP---FHPSPSSPFRPPFNI---KHPFAPNIMPF 212
QY 181 NVGDIFFQ---STQVGPAPFFFTTPPESTF REAFVIFVGVVHVPFELSECHIFF 216
D1 181 NVGDIFFQ---STQVGPAPFFFTTPPESTF REAFVIFVGVVHVPFELSECHIFF 216
QY 233 GYGLNLNLEMEHFFPMI-HUAGGAMVNLHPLHHEIMFFELHAGQSAV KEIHNSTG 291
D1 233 GYGLNLNLEMEHFFPMI-HUAGGAMVNLHPLHHEIMFFELHAGQSAV KEIHNSTG 291
QY 237 QHPHGHRLPFLQFEMTQHMLOGGHGANEHPLGGGATF-SNFATTPAV PEIHNNA 295
D1 237 QHPHGHRLPFLQFEMTQHMLOGGHGANEHPLGGGATF-SNFATTPAV PEIHNNA 295
QY 295 TTPWYTSFTFTTITVWTSQDAVAVLGLNLNLEMEHFFELHAGQSAV KEIHNSTG 326
D1 295 TTPWYTSFTFTTITVWTSQDAVAVLGLNLNLEMEHFFELHAGQSAV KEIHNSTG 326
QY 296 CLEMEDCEKCPREILAVDCQCTFVSGLPFEPEDALPLAEPTF 340
D1 296 CLEMEDCEKCPREILAVDCQCTFVSGLPFEPEDALPLAEPTF 340

RESULT 10
150131
clusterin - quat
Accession: 150131
C:Date: 13-Sep-1996 #sequence revision 13 Sep 1996 #next change 13 Aug 1999
R:Michel, D.; Chacelain, S.; Herault, Y.; Brun, G.
Eur. J. Biochem. 229, 215-223, 1995
Article: the expression of the avian clusterin gene has been cloned, two alternative promoters identified, and the structure of the gene determined.
A:Accession: 150131
A:Reference: 150131, NCBI:55262678; PMID:77441
A:Status: preliminary; translated from cDNA
A:Molecule type: DNA
A:Residues: 1451; <MC>
A:Cross-references: EMBL:X80740; NID:946299; PIR:TAAG033.1; FID:42940
A:Gene: T44
A:Superfamily: clusterin


```

DB 1149 VLCD LDDDEUINOGTSTGCVGQVLEKLEASVSGSELESEETETETWELPNDI 1206
CY      RUMREKELLINLEKAPPELEKALPILKEMPLA LUMH IMAWNE YI TPT 119
DB 1207 SGLQGLMMLCADCASERQQLLEFVLEKLEK-----MFTPEASPYK-- 1255
CY 120 CMFFAFVGLGGLGUGVVEEINLJFFENUNDEECLLENECCHALZACDSE 179
DB 1254 LLEEVRECC LEEENNE
CY 180 DRASSIMD 187
DB 1301 KMBEUTE 1308

RESULT 14
US-10-153-064-73
; Genebank 73, Application 05/10/1994
; Patent No. US20020142814A1
; GENERAL INFORMATION:
; APPLICANT: Bell et al.
; TITLE OF INVENTION: Chemokine beta 1 fusion proteins
; FILE REFERENCE: PFS56
; CURRENT APPLICATION NUMBER: US20020142814
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 00/022746
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 73
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-064-73

Query Match
Best Local Similarity 95.7%, E=1.0e-16, Length 22,
Matches 18, Conservative 2, Mismatches 1, Indels 0, Gaps 0

CY 1 METLLGGLTLLTWENKRLS 21
DB 2 METLLGGLTLLTWENKRLS 22

RESULT 14
US-10-108-605-213
; Sequence 213, Application US/01/012903
; Patent No. US20020142814A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Ravi
; TITLE OF INVENTION: NOVELLACTIC ESTERES AS INHIBITORS OF HIV-1 REPLICATION
; TITLE OF INVENTION: PEPTIDE ESSENTIAL FOR TACTILE VIABILITY AND OTHER THEREOF
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/01/012903
; PRIOR FILING DATE: 2001-03-17
; PRIOR APPLICATION NUMBER: US/97/06114
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 213
; LENGTH: 1374
; TYPE: PRT
; ORGANISM: Grosophila melanogaster
US-10-108-605-213

Query Match
Best Local Similarity 17.1%, E=1.0e-01, Length 1374,
Matches 67, Conservative 71, Mismatches 140, Indels 114, Gaps 17;

```

```

CY 23 QMVSDELEEMSTEGSYNFEELNKLKVV-----KQIKTLEKTN 44
DB 043 ELKQBELQCTGSHACQVEAFAPALFNNVQCHTTPSSQCHPTANFQW 41
CY 54 EEPFLITLCEEEEREP FALADPDSPPQ FADQDQ QMMAW 103
DB 54 EEPFLITLCEEEEREP FALADPDSPPQ FADQDQ QMMAW 103
CY 110 EEL KCLTNTPEVAFVETCTT VVCTFFHMAQPTVFWNPTPTCTT 100
DB 110 EEL KCLTNTPEVAFVETCTT VVCTFFHMAQPTVFWNPTPTCTT 100
CY 363 LEADSGDELEGEQHONLESSQAYDQKLE-----KELTALSNSSA 407
DB 363 LEADSGDELEGEQHONLESSQAYDQKLE-----KELTALSNSSA 407
CY 167 CQTHALLVMGSHSPASSIMRETPQPTREACQPHESPPSSPQ-- 414
DB 167 CQTHALLVMGSHSPASSIMRETPQPTREACQPHESPPSSPQ-- 414
CY 408 ITRGHITAVANILCA----- 464
DB 408 ITRGHITAVANILCA----- 464
CY 515 EFERIPHEEA FUMPEPTVATNFEDMTPTPTVTHQAC QANVNVHRIHPH 469
DB 515 EFERIPHEEA FUMPEPTVATNFEDMTPTPTVTHQAC QANVNVHRIHPH 469
CY 446 QVTEFVAVVMSSTANFE-GENVE FEARILKAMQVQONRQOOF 501
DB 446 QVTEFVAVVMSSTANFE-GENVE FEARILKAMQVQONRQOOF 501
CY 260 MEETFEENQGANHELEINNETLLENECTPTPTTTPMOQSN 519
DB 260 MEETFEENQGANHELEINNETLLENECTPTPTTTPMOQSN 519
CY 500 MARETTEFEELAMVAFARCTPTPTVAVAVV RAEQVQVEMPTLLA PLINIA 509
DB 500 MARETTEFEELAMVAFARCTPTPTVAVAVV RAEQVQVEMPTLLA PLINIA 509
CY 115 EEFEMNDCGICAFPTPTVQVRA AAVLPG 347
DB 115 EEFEMNDCGICAFPTPTVQVRA AAVLPG 347
CY 509 ---PELQSEGEFCQRTAVIRGVHDIASMLPG 587
DB 509 ---PELQSEGEFCQRTAVIRGVHDIASMLPG 587

RESULT 15
US-10-118-513A-12
; Genebank 12, Application 05/10/1994
; Publication No. US20020030955A1
; GENERAL INFORMATION:
; APPLICANT: Kimura, Naoki
; TITLE OF INVENTION: THE 100% NEW INVENTED IN FEMITIVE HEMATOPOIESIS
; FILE REFERENCE: 06501-107US1
; CURRENT APPLICATION NUMBER: 05/10/118,513A
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION: 97-04988 10/07/1997 065756
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: JP 11 288738
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: JP 11 288739
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: JP 2000 123721
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 12
; LENGTH: 2243
; TYPE: PRT
; ORGANISM: Mif musculus
US-10-118-513A-12

Query Match
Best Local Similarity 31.6%, E=1.0e-12,
Matches 86, Conservative 48, Mismatches 159, Indels 96, Gaps 24;

```

```

CY 4 MELVEJLQWVLEHLEKLEPTPTTP FIPALFVPTVPTPTPTPTPT 66
DB 4 MELVEJLQWVLEHLEKLEPTPTTP FIPALFVPTVPTPTPTPTPT 66
CY 511 VESL ELLEHPTATPAWHEPITPAV-LEHVALEKLEKLEKLEKLEK 403
DB 511 VESL ELLEHPTATPAWHEPITPAV-LEHVALEKLEKLEKLEKLEK 403
CY 50 EELITN LEEAFERELANTTPTTENVIRATVTTTVAIREEVQPC-- 110
DB 50 EELITN LEEAFERELANTTPTTENVIRATVTTTVAIREEVQPC-- 110
CY 838 LTVLLFQNVKAWHLRQNRHWEEL LKHAHVQCE MELVELLKEPFTTEG 446
DB 838 LTVLLFQNVKAWHLRQNRHWEEL LKHAHVQCE MELVELLKEPFTTEG 446
CY 100 CMKTVARYRCGTOLVQHO VPELQCNFVFWINGRIDSLRNDQTHALVMS 107
DB 100 CMKTVARYRCGTOLVQHO VPELQCNFVFWINGRIDSLRNDQTHALVMS 107
CY 300 EL VPELLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 103
DB 300 EL VPELLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 103

```


The first of these is the fact that the
 government has been unable to
 maintain a stable currency. This
 has led to a loss of confidence
 in the government and a
 consequent loss of support
 from the people. The second
 is the fact that the government
 has been unable to maintain
 a stable economy. This has
 led to a loss of confidence
 in the government and a
 consequent loss of support
 from the people. The third
 is the fact that the government
 has been unable to maintain
 a stable society. This has
 led to a loss of confidence
 in the government and a
 consequent loss of support
 from the people.

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

XX 17-OCT-2000; 2000W0 US28644
 XX 02-NOV-1999; 99NS-0163685
 PR 17-DEC-1999; 99NS-0172411.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
 PI Komatsoulis G, Poser CA, Fokko SM, Juan PC, Young PE, Floriano KA;
 PI Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
 XX WPI; 2001-328782/34.
 DR N-PSUB; AAD08192
 XX Novel human secreted proteins and nucleic acids for diagnosing,
 PT preventing and treating neurological, cardiovascular, infectious,
 PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
 PT ovarian cancer
 XX Claim 11: Page 387-390; 421pp; English.
 XX AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
 CC protein genes and AAK03764-AAK03786 represent the proteins they encode.
 CC AAK03787-AAK04800 represent human secreted protein fragments or variants.
 CC The genes and their secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 19 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumors, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, androgenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infectious. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention.
 XX SQ Sequence 766 AA;

Query Match 100.0%, Score 4081; Dr 22; Length 766;
 Best local Similarity 100.0%; pred No. 0;
 Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWRPAGAELESLMALWENTALSTWVAVAAVSTLQIATSPFELWLLSLKGPHPSPQFY 60
 DB 1 MWRPAGAELESLMALWVAVAAVSTLQIATSPFELWLLSLKGPHPSPQFY 60
 QY 61 TQVDRSPQCFSTFYKTYPEFGKVNNI AVEPRNPLASPLAPPEFPNPLIPPLPPTI 120
 DB 61 TQVDRSPQCFSTFYKTYPEFGKVNNI AVEPRNPLASPLAPPEFPNPLIPPLPPTI 120
 QY 121 QQTENLKKYGHLLDZADZEPSTLTPVAFPLSPAPGSSSTNSSVPTFHQ 180
 DB 121 QQTENLKKYGHLLDZADZEPSTLTPVAFPLSPAPGSSSTNSSVPTFHQ 180
 QY 181 AASYFDGKSTLPPHLLQIATATKVTFTPTGPGTGSNYNINLGSVSVLVSPENKIQ 240
 DB 181 AASYFDGKSTLPPHLLQIATATKVTFTPTGPGTGSNYNINLGSVSVLVSPENKIQ 240

QY 241 AGTQVLLPVTQPRFVQAAI SVTACNSPPTCKFNKFWHCTIPEPPTCSMTQIAME 300
 DB 241 AGTQVLLPVTQPRFVQAAI SVTACNSPPTCKFNKFWHCTIPEPPTCSMTQIAME 300
 QY 301 PNLIRITETWKAAYNSDPFSPKPKPKPMNYVEI NESTIMHIWIMDSNPPPPYQLEN 360
 DB 301 PNLIRITETWKAAYNSDPFSPKPKPKPMNYVEI NESTIMHIWIMDSNPPPPYQLEN 360
 QY 361 SNKQLFLKAKIVIKKLSLSPKCHKOPLISLPPTTYWLTPTQSTLYCNEWILLSEFS 420
 DB 361 SNKQLFLKAKIVIKKLSLSPKCHKOPLISLPPTTYWLTPTQSTLYCNEWILLSEFS 420
 QY 421 PPTHSTQTPNQVSTAPFPCITVSTPASACTACATVSTPSTPSTPSTPSTPSTPSTPST 480
 DB 421 PPTHSTQTPNQVSTAPFPCITVSTPASACTACATVSTPSTPSTPSTPSTPSTPSTPST 480
 QY 481 STDHYIGFETDLQDLEMKYLLQKTDORRFEVHAIFISNMKINSMWFDPSWKKPMILILKSN 540
 DB 481 STDHYIGFETDLQDLEMKYLLQKTDORRFEVHAIFISNMKINSMWFDPSWKKPMILILKSN 540
 QY 541 KYKSSLVEMILGLSLQICITKNTSLPEVLAVVNPFGSSHSFMPVNPVNSPDPWEPTK 600
 DB 541 KYKSSLVEMILGLSLQICITKNTSLPEVLAVVNPFGSSHSFMPVNPVNSPDPWEPTK 600
 QY 601 LQPLQCTYNTWTLTGKWKKEFFVHYIYKSPKSNQPNQSPSYVEKPEFIDPSNLYG 660
 DB 601 LQPLQCTYNTWTLTGKWKKEFFVHYIYKSPKSNQPNQSPSYVEKPEFIDPSNLYG 660
 QY 661 MKINNTQFGYSMMRDPKAIHLLIQLQVYFGSGSALQILEHRYVKNKLSIPQDQR 720
 DB 661 MKINNTQFGYSMMRDPKAIHLLIQLQVYFGSGSALQILEHRYVKNKLSIPQDQR 720
 QY 721 IDLFSCLLPHLPKLTSEVVRITQSAIQAFNAKLENTMDYDTIKLS 766
 DB 721 IDLFSCLLPHLPKLTSEVVRITQSAIQAFNAKLENTMDYDTIKLS 766
 RESULT 2
 AAG65480
 ID AAG65480 standard; Protein; 766 AA.
 XX AC AAG65480;
 XX DT 27-AUG-2002 (first entry)
 XX Human albumin fusion protein #2155.
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA, cancer, reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antiinfertility; antinflammatory; antiulcer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic
 OS Homo sapiens.
 OS Synthetic.
 XX WO200177137-A1.
 XX 18-OCT-2001.
 XX 12-APR-2001; 2001WO-US11988.
 XX 12-APR-2001; 2000US-224358P.
 PR 25-APR-2000; 2000US-199384P.
 PR 21-DEC-2002; 2000US-256931P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Poser CA, Haseltine WA;
 PI

hyperproliferative disorder, sleep pattern, cardiovascular disorder;
reproductive disorder, digestive system disorder, behavioural disorder;
Homo sapiens.

W0200216390 A1.

28 FEB-2002.

17 JAN-2001; 2001WO-0501435.

18 AUG-2000; 2000US-024362P.

(HUMA) HUMAN GENOME SCI INC.

Rosen CA, Komatsoulis CA, Baker KP, Hirse CH, Soppet DR, Olson HS;
Moore PA, Wei P, Ebner K, Duan DR, Shi Y, Choi GH, Piscella M;
NI J;

WPI; 2002-304113/34.

An isolated nucleic acid molecule (1) comprising a polynucleotide which
encodes a polypeptide useful in the diagnosis and treatment of
disorders e.g. immune disorders -

Disclosure: Page 59; 534pp; English.

AA033692-AA033736 represent cDNAs corresponding to 21 human secreted
protein genes, and AA021191-AA021235 represent the proteins they encode.
AA021236-AA021280 represent human secreted protein fragments. The genes
and their corresponding secreted proteins are useful for preventing,
treating or ameliorating medical conditions, e.g., by protein or gene
therapy. Pathological conditions can be diagnosed by determining the
amount of the new protein in a sample or by determining the presence of
mutations in the new genes. Specific uses are described for each of the
21 genes, based on the tissues in which they are most highly expressed,
and include developing products for the diagnosis or treatment of
immune or autoimmune diseases e.g. AIDS (acquired immune deficiency
syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
and breast cancer, neurological diseases e.g. Alzheimer's disease,
Parkinson's disease, Huntington's disease, Tourette syndrome,
meningitis, demyelinating disease, peripheral neuropathies, toxic
trauma, congenital malformations, spinal cord injuries, multiple
neuropathies induced by neurotoxins, peripheral neuropathies, multiple
sclerosis, ischaemia and infarction, haemorrhages, schizophrenia, mania,
dementia, depression, panic disorder, learning disabilities, AIDS,
altered behaviours e.g. disorders in feeding, sleep patterns, balance
and perception, encephalitis, disorders in cardiovascular, neural/
sensory, reproductive and digestive systems, behavioural disorders and
hyperproliferative disorder. The present sequence represents human
secreted protein fragment referred to in the disclosure of the invention.

Sequence 347 AA:

Query Match 41.4%; Score 1691; Dn 24; Length 337;
Best local Similarity 100.0%; Pred. No. 1,26-139;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MIWRSAGAEHLSMALWMIWIALSLHGWVAVAVSDQHATSPEDWLLSDKGPFRSQEY 60

1 MIWRSAGAEHLSMALWMIWIALSLHGWVAVAVSDQHATSPEDWLLSDKGPFRSQEY 60

61 TDVDSRQGFSTRYKLYEFGHKKVNNIADVPRNPLGSPPLAPFPFNNPLGPPPPH 120

61 TDVDSRQGFSTRYKLYEFGHKKVNNIADVPRNPLGSPPLAPFPFNNPLGPPPPH 120

121 QQTENI IKKYGTHFLSATLGGFSSITFVDFKFLSKPAFGSDSTNSSSVTLLELHQL 180

121 QQTENI IKKYGTHFLSATLGGFSSITFVDFKFLSKPAFGSDSTNSSSVTLLELHQL 180

181 AASYFIDRUSTLPRHHIOASTAKVETRTGPGCCSNYNDLSVSVSVSPENKIQI 240

181 AASYFIDRUSTLPRHHIOASTAKVETRTGPGCCSNYNDLSVSVSVSPENKIQI 240

QY 241 QSLQVLLFTYLLFFVQAALSYIAVNSSEETLEHNSWELASFLLETRHLMGLVAML 400
DB 241 QSLQVLLFTYLLFFVQAALSYIAVNSSEETLEHNSWELASFLLETRHLMGLVAML 400

QY 301 ENLLRITETWKAYNSDFEES 320

DB 301 ENLLRITETWKAYNSDFEES 320

RESULT 7

AAM58347

ID AAM58347 standard: Protein; 478 AA.

XX AAM58347;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 40452.

XX Human; brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer.

XX Homo sapiens.

XX WC200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-0500667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 04-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0034263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains.

XX Example 4; SEQ ID NO: 40452; 650pp - Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is a protein encoded by one of

XX the probes of the invention.

XX Sequence 378 AA:

Query Match 38.0%; Score 1550; Dn 22; Length 378;

Best local Similarity 74.7%; Pred. No. 1,5-127;

Matches 278; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 395 RTSTWLTQISFLYCNENGLISFSEETHSCIPNQQVVGCTAFPLPTVSDASALITAP 474

DB 7 RSLVWNNRIQLLYGSGESTFGTELESHSCTFYDSSSQGEIPCALGEGPAWACAP 66

QY 455 DNRTRCCGTCNTGYMLSSQGLCKHVEANSTDRYIGETHLQGLEMAVLLQKTRDRRIVHAF 514

DB 67 DNSTRTGSGNPGYVLAAGLCPPEVAESLENLEGLTELEQLQLELYLLQKQDSFEVHSIF 126

PR 03-AUG-2000; 2000US-0632466.
 PR 21-SEP-2000; 2000US-0234487.
 PR 27-SEP-2000; 2000US-0234459.
 PR 04-OCT-2000; 2000GB-0C24263.
 XX (MOLFE-) MOLECULAR DYNAMICS INC.
 XX PA Penn SG, Hanzel DR, Chen W, Rank DR;
 XX PI WPI; 2001 483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PI brains -
 PS Example 4: SEQ ID NO: 30424; 650pp - Sequence Listing: English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX Seq Sequence 470 AA:
 Query Match 26 Pk Score 1097.6; FR 22; Length 470;
 Best Local Similarity 60 Pk Prod No 2 7e-91;
 Matches 10; Conserved 10; Mismatch 10; Index 10; Gap 0.
 QY 396 TSYWLTIRUSLYCNGHNGHSGFSETHSCDNDVDDVDTAFCTVDDASATCTADP 455
 Db 1 TIOGTLARVUSLLYCNENFMTFTESQSSVHRSSTTHGPPETVGNNSVAMSLA 66
 QY 456 NPTQSTNTSYMLSGGKEEVAES-ATHYVGEETD-GLLEMKYLLAKILKPTIEVH 511
 Db 61 NISGSCNKGYLEYRPPPPNVNVSPPSQFSEFDHDEDDPKYVTKRMSPFVHH 120
 QY 512 APTSNMPLNSWPDSSWPPMELITPSNPFYSSIVMTGTEGLQGLTPTSTTEFFVLAV 571
 Db 121 TPTSNTEITDTEPPWFFPMSTITKSNKNMPTTHVVMSPICOMPNSSTFMEFF 180
 QY 572 VVNPESSSSSWFMVNSNVAWPTKPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 631
 Db 181 VVNPESSSSSWFMVNSNVAWPTKPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 240
 QY 632 RTRSNPNNESTVYPTETPLSPALNMYKINNNVAVSYSMPEPPPEAPLPLGLLQVY 691
 Db 240 RTRPTLTPNET-CCGQDVDDSHSKKQPVYKISDQVQVYSPENADLPSAVGVWQSY 298
 QY 692 TQNSQ-----DSALLQLLTPDPVNRISP---PQGPPLNPSGLPIRLPTSTSPVVPVQS 744
 Db 299 TQNSQFYSSSSVMLELTPETNELAPPVAPKPKQLLESIMIKRLKLNSELIPVNH 358
 QY 745 ALQAFNAKIPNIMVYDTKE 765
 Db 359 ALDLYNTEILKQSDMTAKLE 474
 RESULT 10
 AHH14717
 XX AHH14717 standard; protein; 121 AA
 AC ABB14717;
 XX 23 JAN 2002 (first entry)
 XX Human nervous system related polypeptide. Seq ID No 3374
 DE Human; neurotropic; neuroprotective; cytoprotective; dermatological; virucide;
 KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
 KW antiparkinsonian; antiskinning; antianemic; antithyroid; cancer;
 KW antirheumatic; hepatotropic; corobroprotective; anti-inflammatory;
 KW

KW anti-allergic; antidiabetic; anti-leucor; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotoxic; gene therapy; vaccine;
 XX Homo sapiens.
 XX WC200159063-A2.
 XX 16-AUG-2001.
 XX 17-JAN-2001; 2001WO-US01334.
 XX 31-JAN-2000; 2000US-0179065
 PR 04-SEP-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186550.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0192123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0225779.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 24-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0227009.
 PR 01-SEP-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.

XX OS Homo sapiens.
 XX PN WO200216390-A1
 XX PD 28-FEB-2002.
 XX PF 17-JAN-2001; 2001WO 9501435.
 XX PP 16-AUG-2000; 2000US 225-82P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 XX PJ Moore PA, Wei P, Edner R, Duan DR, Shi Y, Choi GH, Fiscella M;
 XX PJ Ni J;
 XX DR WPI: 2002-304113/34.
 XX PT An isolated nucleic acid molecule (I) comprising a polynucleotide which
 XX PT encodes a polypeptide useful in the diagnosis and treatment of
 XX PT disorders e.g. immune disorders.
 XX PS Disclosure; Page 59; 534pp; English.
 XX CC AAD33692-AAD33736 represent cDNAs corresponding to 21 human secreted
 XX CC protein genes, and AAE21191-AAE21235 represent the proteins they encode.
 XX CC AAE21236-AAE21280 represent human secreted protein fragments. The genes
 XX CC and their corresponding secreted proteins are useful for prevention,
 XX CC treating or ameliorating medical conditions, e.g., by protein or gene
 XX CC therapy. Pathological conditions can be diagnosed by determining the
 XX CC amount of the new protein in a sample or by determining the presence of
 XX CC mutations in the new genes. Specific uses are described for each of the
 XX CC 21 genes, based on the tissues in which they are most highly expressed,
 XX CC and include developing products for the diagnosis or treatment of
 XX CC immune or autoimmune diseases e.g. AIDS (acquired immune deficiency
 XX CC syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
 XX CC and breast cancer, neurological diseases e.g. Alzheimer's disease,
 XX CC Parkinson's disease, Huntington's disease, Tourette syndrome,
 XX CC meningitis, demyelinating disease, peripheral neuropathies, neoplasia,
 XX CC trauma, congenital malformations, spinal cord injuries, toxic
 XX CC neuropathies induced by neurotoxins, peripheral neuropathies, multiple
 XX CC sclerosis, ischaemia and infarction, haemorrhages, schizophrenia, mania,
 XX CC dementia, depression, panic disorder, learning disabilities, ALS,
 XX CC altered behaviours e.g. disorders in feeding, sleep patterns, balance
 XX CC and perception, encephalitis, disorders in cardiovascular, neural/
 XX CC sensory, reproductive and digestive systems, behavioural disorders and
 XX CC hyperproliferative disorder. The present sequence represents human
 XX CC secreted protein fragment referred to in the disclosure of the invention.
 XX SQ Sequence 88 AA;
 Query Match 10.6%; Score 433; DB 23; length 88;
 Best Local Similarity 100.0%; Pred No 4 2e 30;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 144 FESLTFVQKRLSKRAFGSDSTNSSSVTLFTLHQLAASFTFDPGSLPLRLHIQIAST 203
 DB 1 FESLTFVQKRLSKRAFGSDSTNSSSVTLFTLHQLAASFTFDPGSLPLRLHIQIAST 60
 QY 204 AIKVTETRTGPGCCSNYDNLDSVSSVLV 231
 DB 61 AIKVTETRTGPGCCSNYDNLDSVSSVLV 88
 RESULT 12
 ABG64856
 ID ABG64856 standard; Protein: 76 AA.
 XX AC ABG64856;
 XX DT 27-AUG-2002 (first entry)
 XX

DE Human albumin fusion protein #1531.
 XX OS Homo sapiens.
 XX KW Albumin fusion protein, therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA, cancer, reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antiinfertility; antiinflammatory; antiulcer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurologic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 XX OS Osteopathic, antiarthritic.
 OS Homo sapiens.
 OS Synthetic.
 PN WO200177137-A1.
 XX PD 18-OCT-2001.
 XX PF 12-APP-2001; 2001WO-US11988.
 XX PP 12-APP-2000; 2000US-229154P
 XX PR 25-APR-2000; 2000US-199384P.
 XX PR 21-DEC-2000; 2000US-256931P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Haseltine WA;
 XX PJ Wei; 2002-010629/01.
 XX PT New fusion protein for treating disease e.g. diabetes comprises an
 XX PT albumin fused to a therapeutic protein.
 XX PS Claim 1, Page 1551, 2102pp, English.
 XX CC The present invention relates to albumin fusion proteins comprising a
 XX CC therapeutic protein X and human albumin (HA), also known as human serum
 XX CC albumin, HSA). The proteins are useful for treating a disease or
 XX CC disorder that may be modulated by therapeutic protein X. The albumin
 XX CC extends the shelf life of protein X, and may increase its biological
 XX CC in vitro/in vivo activity. The protein is useful for treating and
 XX CC diagnosing disorders such as cancer, reproductive disorders, digestive
 XX CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 XX CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 XX CC (e.g. diabetes), haematopoietic disorders, neural disorders
 XX CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 XX CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 XX CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
 XX CC fusion proteins of the invention.
 XX SQ Sequence 76 AA;
 Query Match 9.9%; Score 403; DB 23; length 76;
 Best Local Similarity 97.4%; Pred No 1 4e 27;
 Matches 74; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MIWRKACAEPLFSIMALMHWIALSIHCWVI AAVASVQHATSPFDWLSKPKGPHRSQRY 60
 DB 1 MIWRKACAEPLFSIMALMHWIALSIHCWVI AAVASVQHATSPFDWLSKPKGPHRSQRY 60
 QY 61 IDFVQKSKGSESTYK 76
 DB 61 IDFVQKSKGSESTYK 76
 RESULT 13
 AAE21211
 ID AAE21211 standard; Protein: 76 AA.
 XX AC AAE21211;
 XX DT 01-JUL-2002 (first entry)
 XX

Query Match	9.5%	Score 403	IR 24	Length 76
Best Local Similarity	97.4%	Prod. No. 1.4e+27		
Matches 74	Conservative 0	Mismatch 2	Indels 0	Gaps 0
QY 1	MWRSPAGAEFESTMALWFWATLSHCWJAVAAVSIQAHSNP	WLLSRQGFHSRQRY 60		
DD 1	MWRSPAGAEFESTMALWFWATLSHCWJAVAAVSIQAHSNP	WLLSRQGFHSRQRY 60		
QY 61	TDFVDRSRQGFSTRYK 76			
DB 61	TDFVDRSRQGFSTRYK 76			

RESULT 14
AAE21276
ID AAE21276 standard; peptide: 74 AA.
XX
XX
AC AAE21276;
XX
XX
DT 01-JUL-2002 (first entry)
XX
DE Human gene 21 encoded secreted protein fragment. SPO ID: N:142.

Human; secreted protein; immune disorder; antiallergic; antirheumatic
 KW
 Rheumatoid arthritis; breast neoplasia; breast cancer; antirheumatic
 KW
 Neurological disease; Alzheimer's disease; Parkinson's disease; brain
 KW
 Tourette syndrome; encephalitis; cysticercosis; basophilic; adenoma; mu
 KW
 Antinflammatory; ophthalmological; dermatological; immunosuppressi
 KW
 Immunomodulatory; immunosuppressive; antidiabetic; antidiabetic
 KW
 Gene therapy; autoimmune disease; bone marrow; disease; immunizat
 KW

XX	Homo sapiens.
XX	reproductive disorder; digestive system disorder; behavioral disorder
KW	hyperplastic disorder; sleep pattern; cardiovascular disorder
KW	depression; panic disorder; learning disability; AIDS; feeding disorder
KW	multiple sclerosis; infarction; hemorrhage; schizophrenia; dementia
KW	spinal cord injury; peripheral neuropathy; ischemia; perception

XX Homo sapiens.
 XX OS
 XX W0200216340-A1.
 XX PN
 XX 28-FEB-2002.
 XX PD
 XX 17-JAN-2001; 2001W00US014+5.
 XX PE

15-ACG-2000, 15-ACG-2526AP.

(HUMA-) HUMAN GENOME: SCI INC.

Rosen CA, Komatsu S, Baker KP, Birse CE, Sorpet DR, Olsen
Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M,
Ni J;
WPI: 2002-04113/34.

An isolated nucleic acid molecule (1) comprising a polynucleotide wh
encodes a polypeptide useful in the diagnosis and treatment of
disorders e.g. immune disorders .

Disclosure: Page 58-59; 534pp: English.

AA033692-AA033746 represent cDNAs corresponding to 25 human secreted
protein genes, and AA031191-AA031210 represent the fragments they enco
AA031236-AA031280 represent human secreted protein fragments. The que
and their corresponding secreted proteins are useful for preventing, trea
treating or ameliorating medical conditions, e.g., by protein or gene
therapy. Pathological conditions can be diagnosed by determining the co
amount of the new protein in a sample or by determining the presence

CC mutations in the new genes. Specific uses are described for each of the
 CC 21 genes based on the tissues in which they are most highly expressed.
 CC and include developing products for the diagnosis or treatment of
 CC immune or autoimmune diseases e.g. AIDS (acquired immune deficiency
 CC syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
 CC and breast cancer, neurological diseases e.g. Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease, Tourette syndrome,
 CC meningitis, demyelinating disease, peripheral neuropathies, neoplasia,
 CC trauma, congenital malformations, spinal cord injuries, toxic
 CC neuropathies induced by neurotoxins, peripheral neuropathies, multiple
 CC sclerosis, ischaemia and infarction, haemorrhages, schizophrenia, mania,
 CC dementia, depression, panic disorder, learning disabilities, AIDS,
 CC altered behaviours e.g. disorders in feeding, sleep patterns, balance
 CC and perception, encephalitis, disorders in cardiovascular, neural/
 CC sensory, reproductive and digestive systems, behavioural disorders and
 CC hyperproliferative disorder. The present sequence represents human
 CC secreted protein fragment referred to in the disclosure of the invention.

XX SQ Sequence 73 AA:

Query Match 9.7%; Score 396; DB 24; Length 73;
 Best Local Similarity 100.0%; Pred. No. 5,56-27;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIMKSRACAEFLSLMALWEIALLSLHCWVLAVAASVDQHATSPFDWLLSKGPFHRSQFY 60
 Db 1 MIMKSRACAEFLSLMALWEIALLSLHCWVLAVAASVDQHATSPFDWLLSKGPFHRSQFY 60

QY 61 TDFVDRSRKGFSST 73

Db 61 TDFVDRSRKGFSST 73

RESULT 15

ABB42190

ID ABB42190 standard; Peptide; 78 AA.

XX ABB42190;

XX 01-FEB-2002 (first entry)

XX Peptide #4841 encoded by breast cell single exon nucleic acid probe.

XX Human: microarray; single exon probe; gene expression; breast;
 XX disease; cancer.

XX Homo sapiens

XX W0200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001W0-0500562.

XX 04-FEB-2000; 2000US-0180312.

XX 24-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0604048.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0024263.

XX (MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001 496933/54.

XX New spatially addressable set of single exon nucleic acid probes,
 XX useful for measuring gene expression in sample derived from human
 XX breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 15158; 327bp; sequence listing; English.

XX the invention relates to a spatially addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or diagnosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from Wipo at ftp.wipo.int/pub/published_pat_sequences.

XX SQ Sequence 78 AA:

Query Match 8.0%; Score 327; DB 22; Length 78;

Best Local Similarity 70.5%; Pred. No. 76-21;

Matches 55; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 243 LOVLLPDYLOEPFVQAALSYTACNSSEGEFTCKENQWVDCSPKTFECWTFPSMDTQAMEEN 402

Db 1 LOVLLPDYLRERFVAALSYITCSSHGLVCKENQWVDCSPKTFECWTFPSMDTQAMEEN 60

QY 303 LDRITETWKAYNSDFEES 320

Db 61 LIQIQDSWATHNRQPEES 78

Search completed: May 19, 2003, 10:31:18

Job time : 83 secs

QY 482 TDHYHFFTLQDL-----FMKVLKQDSEIEVHAIFISNDWPLNSWDSWKRKML 534
 DB 521 RNYTKKMSINDIYVRAISEHLYFYSKYKERV---LFWNNM--PNYNGSKLRF 574
 QY 535 LTKSNKYKSSIVHMIHGL-----SIQICHTK 561
 DB 575 LEVNSDFQ-NITPITGLSLSYHNPALYYKDIIKIPETWNRKSEAKTILDSLKNWLT- 632
 QY 562 NSTLEIVL---AVYVNYFCGSUSHSZWPVNVINSP FHWERTKLEL 603
 DB 633 NNNVDFVRWKYTDQIYIYID SPDITLWLFOSIKRHSFOLLNVHVSCTNTEILELV 692
 QY 604 PL-----QCYNWLTL-----GNKWKTEFEVHI YLR- 636
 DB 693 PETHSVNAHQIYVYACNIYTLFDSGSKIKYFYAKIYVNRKOPOTSPLOKEYPILLIKYIKL 752
 QY 631 -----SRIKNGVNGNE-----SIYVEHFDIDPSKNIQYKNNIQVFC 670
 DB 753 PENIQWYIYRKDKRSTELRTPYVFDKQVGLYSRFPFWATKAP-----KNIQIS 805
 QY 671 YSM-HEDPEAIR-----DLILQIYIYPTGCSOVSALIGLEHNR 709
 DB 806 YEISEYIPESARYCGKXKYSIHSFFYYUSILDLILQ-----KKARQIEVTS 854
 QY 710 VN-----KLSPPQRL--DLFSCLLRHRLK 733
 DB 855 LSTIRMLMKLSIQAQKLIKRNKIFAG--KKRLK 886

RESULT 2

US-09-183-861-22
 ? Sequence 22, Application 05/09183861
 ? Patent No. 6365165
 ? GENERAL INFORMATION:
 ? APPLICANT: Reed, Steven G.
 ? APPLICANT: Campos-Neto, Antonio
 ? APPLICANT: Webb, John R.
 ? APPLICANT: Dillon, David C.
 ? APPLICANT: Skeiky, Yasir A.W.
 ? TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
 ? NUMBER OF SEQUENCES: 87
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: SEED and BERRY LLP
 ? STREET: 6300 Columbia Center, 701 Fifth Avenue
 ? CITY: Seattle
 ? STATE: Washington
 ? COUNTRY: USA
 ? ZIP: 98104-7092
 ? COMPUTER READABLE FORM:
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent in Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/09/183,861
 ? FILING DATE:
 ? CLASSIFICATION:
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 09/022,765
 ? FILING DATE: 12-FEB-1998
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Maki, David J.
 ? REGISTRATION NUMBER: 31,392
 ? REFERENCE/DOCKET NUMBER: 210121.420C3
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (206) 622-4900
 ? TELEFAX: (206) 682-6031
 ? INFORMATION FOR SEQ ID NO: 22:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 320 amino acids
 ? TYPE: amino acid
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein

US-09-183-861-22
 Query Match 2.8%; Score 114.5; DR 4; Length 320;
 Best local Similarity 32.3%; Pred. No. 0.02;
 Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;
 QY 383 GHKQPLISLPQRTSTYWLRIQSLYFNENGLGSFSEETHSCTCPNDQVVTAFPLPT 442
 DB 176 FRLSIACSVPNKKKGLGISHL-----CAGDQGYSLSAIAISSSPTQ----- 224
 QY 443 VRIASACITAFNPTRPNTGNTNGYMLSGLEK 475
 DB 224 V-----EHCNTGVNSJDSRCAVNTGYVVSDEKCK 253
 RESULT 3
 US-09-183-861-55
 ? Sequence 55, Application 05/09183861
 ? Patent No. 6365165
 ? GENERAL INFORMATION:
 ? APPLICANT: Reed, Steven G.
 ? APPLICANT: Campos-Neto, Antonio
 ? APPLICANT: Webb, John R.
 ? APPLICANT: Dillon, David C.
 ? APPLICANT: Skeiky, Yasir A.W.
 ? TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
 ? NUMBER OF SEQUENCES: 87
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: SEED and BERRY LLP
 ? STREET: 6300 Columbia Center, 701 Fifth Avenue
 ? CITY: Seattle
 ? STATE: Washington
 ? COUNTRY: USA
 ? ZIP: 98104-7092
 ? COMPUTER READABLE FORM:
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent in Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/09/183,861
 ? FILING DATE:
 ? CLASSIFICATION:
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 09/022,765
 ? FILING DATE: 12-FEB-1998
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Maki, David J.
 ? REGISTRATION NUMBER: 31,392
 ? REFERENCE/DOCKET NUMBER: 210121.420C3
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (206) 622-4900
 ? TELEFAX: (206) 682-6031
 ? INFORMATION FOR SEQ ID NO: 55:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 320 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS:
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? FRAGMENT TYPE: Internal
 ? US-09-183-861-55

Query Match 2.8%; Score 114.5; DR 4; Length 320;
 Best local Similarity 32.3%; Pred. No. 0.02;
 Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;
 QY 383 GHKQPLISLPQRTSTYWLRIQSLYFNENGLGSFSEETHSCTCPNDQVVTAFPLPT 442
 DB 176 FRLSIACSVPNKKKGLGISHL-----CAGDQGYSLSAIAISSSPTQ----- 224
 QY 443 VRIASACITAFNPTRPNTGNTNGYMLSGLEK 475

AYNSOEE-----SDEKLE MKKPMNYPLNTSTIMHLWTM D 348
CMAIHDKKKHLNUNFEMHLEKTKKSLIFLOINTMOORSTIOE 696
CMKPLAKAKIVHKLESKPKPKKPKPKPK 394
CVLLTMTMLKAEMLV KQVVKLINESPK 748
CUB US/9479441
WAY, KURMIT L.
BERS "ARAWAY, CORALIE A.
ON, NAVIS L.
NE INGENE PRODUKT LIANI
EST 125
DRESS
MAN, HARRY & CUSHMAN
NEW YORK AVENUE, N.Y.
N
FORM:
COPY disk
of compatible
M: PC-DOS/MS-DOS
In Release #1.0, Version #1.25
IN DATA:
PAPER: US/89/179,481
#-D8-1993
445
DATA:
SER: US 87/922,621
JUL-1992
INFORMATION:
PAUL N.
MBER: 16,773
NUMBER: 2607027/WL
INFORMATION:
861-000
422-0444
CUSE
NO: 2:
DISTIN:
ago acids
if
CUBA
581 Score 100, 1b 1 Length 744:
1481 Pred. No. 1:
661 Misat. 661 Indels 128: Gaps 26:
-WKVNNLAVRRHFLQSPDPAAREPKNBQ-----LGRPHLCOIT 124
LWELSPATLRMSYASPTSAVAHELEPENSJHVDNIKPKTYL----- 341
SALGSEFSLIFVUKKLSKAPNSNNSSSVILETHOLAASY 184
SSVLCPEIVATPCSKKQTYNYNSKIINSSFEVI
HQLASTATKVIETRTGELPNSNENLNVSSSVLVQSPENKIQLOQ-- 242
SK -----DLCTEP CFINVPLIKKGCQACPNMIGDGRH 436

OY 243 -LOVLLPDYLORE--VQAALSYIACNSDEPCKENHFWHCKKPEPNC 291
Db 437 CVAVEISEFCONISCVNYCYNHGCHDISPPKQPT-----CTCAFTGNKFLANNET 493
OY 292 ----PNDIQAMFENLRTETWKAYNSDFFPSGPKFEMKPLIMNYELNTSTI 441
Db 494 PLIYKELPRTI---TLSLRDEPNASNAUVNAS- VANVLENIEMHAFELNSLVELLRIS 548
OY 342 -----MLWTMDSNFAPP-----YELENSMKPLEI 571
Db 549 PGAPVLGKPIHUKVSHFKYPRGRLIHYLNQLISAVNEAFIQAPEPRKRSSEARK 608
OY 372 IYKLEKLSKKCKKHLSLSPROFTSTIYWRQSPLEYNENACDSESELENS-ENO 441
Db 609 NVRETPIS-RADVQDSMALN-----LSMLREYFTQ-IRYK- YHLYVSPQD 651
OY 432 QVVCTALFCTVQ-----LASALITCA 454
Db 652 GVTCVS -PCSEGYCHNGGCKKHLPIGPGQ-ICA 682
RESULT 10
US-09-004-838-125
; Sequence 125, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Michelemore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; NUMBER OF SEQUENCES: 140
; CONFERENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.40
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,744
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Elthorn, Gregory P.
; REGISTRATION NUMBER: 38,446
; REFERENCE/DOCKET NUMBER: 024076 07881063S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 125:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1817 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY:
; LOCATION: 1..1817
; OTHER INFORMATION: /note- "RG2S deduced sequence"
US-09-004-838 125
Query Match 2.5%; Score 102; DB 4; Length 1817;
Best local Similarity 18.9%; Pred. No. 5,6;


```

Query Match          2.5%; Score 101.5; DB 5; Length 993,
Best Local Similarity 19.3%; Pred. No. 2.3; 237; Indels 183; Gaps 20;
Matches 119; Conservative 76; Mismatches 76;

QY 169 SSSVTLFLHQLAAS VFIDPSTLEPLHLHIQ-----LASTAIKVTETFTGPL----- 215
DQ 86 SASITLQVLDVAICNLSLWVKKHSSINCOPHFDLONKGVSMVILKWTETQACHYLLFL 145

QY 216 --GCSYNDNLDSV-----SVLVQSPENKIQLOGLQVLLPYLOERFVQAALSYIAGNS 267
DQ 146 QSEATNYTILTSIPNTLTYLTPRPYFKMNDALAVTISVDPERIVE-----VWLVDLS 201

QY 268 PGPFKPKP-----NLCWCHQCHQKPK-----PNCNCHPSM 294
DQ 202 QGE SPYEEFAVVKKEKVIHEIFDTLECCAFNELPEPTELETLIDLTGTFQTLPL 260

QY 295 DQAMENLLPTE-----TWKAYNSDEESDEFKLEKPLPMNYFLNTSTIMHLW 345
DQ 261 FLKVGCEPLWIRKAVHVNHPGCLTWLEKKALEKGNFYEM-----SYSTNRTMIRILF 314

QY 346 TMSNFORR---YEOLENSMKOLFKAOKTIVHLIP-----SISKPKYIKQPLI 389
DQ 315 AFVSSVARNDTGYTSSSKHPSQSALVTIVEKGFINATNSSEYDIDQYEEFESVREK 374

QY 390 SLRPORTSYWLIKRIQSPFYCNGL-----LQSPSEETHS-----CTCPNDQVVCYAF 438
DQ 375 AYPQIRCT--WTESRKSF--PCEQKGLDNGYSISKFCNKHQGEYIFHAENDDDAQFTKM 431

QY 439 LPTVGDASACLICAPDNTRCTNTNYGLS-----QGLCKPEVAESTDIHYIGFE 489
DQ 432 PTNIRKKKPOVIAASASQASC--FSDGYPLSWTWKKCDKSPNCTREHTEGVWN----- 484

QY 490 TDLQLEPMKYLLQKTPPTFVIAIFTSNDMPLENSWFDPSWPKPMILLTK--SNKYFSSLVH 548
DQ 485 -----PKANKV-----PGQWVSS-----TLNNSAATKGLVLR 513

QY 549 MLLGLSGLCLTNS---TLEWLVAVVNPFGCSSESFMVNVNSPDMWPKIKIDPL 605
DQ 514 -----GCAYNSLSTSCETILLSPGPF-----PFLQUNISFATIAWV-L-L 553

QY 606 QYNWTLTIAANKKTFETVHYLRSPKSGNPNFSTY-----YE-PLFIDPSRNLG 659
DQ 554 FIVVITLILCHKKYKQPP--YEQIQMVQVTCSSINERYFYVDFPKYKYLKWEFFPENLE 611

QY 660 YMKINNTQVGRYSMH 674
DQ 612 FAKVLGSAFGKVM 626

RESULT 14
US-08-476-515A 84
: Sequence 84, Application US/08476515A
: Patent No. 6239270
: GENERAL INFORMATION:
: APPLICANT: Akorstrom, Goran
: APPLICANT: Juhlin, Claes
: APPLICANT: Bask, Lars
: APPLICANT: Crumley, Gregg R.
: APPLICANT: Morse, Clarence C.
: APPLICANT: Murray, Edward M.
: APPLICANT: Hjalms, Goran
: TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
: TITLE OF INVENTION: Therapeutic and DNA Encoding Same
: NUMBER OF SEQUENCES: 84
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Martin Savitzky
: STREET: Rhone Poulenc Rorer Inc., 500 Arcola Rd.
: STREET: 3C43,
: CITY: Collegeville
: STATE: PA
: COUNTRY: USA
: ZIP: 19426-0107
: COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: Compaq PC
OPERATING SYSTEM: Windows 95
SOFTWARE: Word 7.0 (Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,515A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,836
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/36,747/00483
FILING DATE: 24-MAY-1994
FELER APPLICATION DATA:
APPLICATION NUMBER: SE 9301764-8
FILING DATE: 24-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Savitzky, Martin
REGISTRATION NUMBER: 29,694
REFERENCE/AGENT NUMBER: A1355D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610 454-3816
TELEFAX: 610-454-3808
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 4654 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08 476-515A-84

Query Match          2.5%; Score 101.5; DB 4; Length 4654;
Best Local Similarity 20.3%; Pred. No. 40;
Matches 107; Conservative 50; Mismatches 183; Indels 187; Gaps 31;

QY 268 PPFCKPKNWCHQCHQKPKPECNCHSMIDAMFNLPITETIWA-----YNSDFPES 420
DQ 1708 DAVFCSNSQAVP--PTPECMCHSEDEFQLEEGIC--IENLWEGHFFWLYLSD 1261

QY 321 DPFLEKMKPLPMNYFL--NTSTIMHLWMSNFGPPYPAIENSMKOLEFLKAKIVHLFS 478
DQ 1262 EHNACVYKTCISYFHCDCNCHRAWLCOR-----GRKQDM-----S 1400

QY 479 LKPPCHVQPIISPPAPSTYW-----LTPYCTSTY-----QSERGLL 419
DQ 1301 DEKQCTPFT--PCTSWQCTLGHNI--VNLAVVLSLLELLENGLHESLGNNSQSD 1455

QY 418 SPSEETHSCH-----CP-----NLCWVCTAFPLCTV-GLASA-GLICAPDNTRC 460
DQ 1356 FNGGCTHECVQEEFRKAKLCPLGLFLLANISKTCEDIDEDLGSFSQHCYNNR--GSEFC 1413

QY 461 GTGNTGYML--SQG--LCKPEVAEST-----CP-----DHYIGFE 489
DQ 1414 -SDTGYMLFSDGPTCKVTASPSILLIYASQNK--IADSVISQVENVYSLVNSCYLVAVD 1472

QY 490 TP-----LQPEMYVLYQYTPPEVVAIFISNEMFLNSWTFP--WPKPMILT 546
DQ 1473 FDSISGRFWSDATQGTWSAFQNGTDRVVFDSSILLITETIAHWVGRNLYWLYVALET 1542

QY 537 LKSNKYKSSLVHMLGLSLQICLTKNSTLEPVLAVVNPFGCSSESFMVNVNSPDM 596
DQ 1549 IEVSK-----IDGSHPTVLISKNLINPECLA--LLEPMNEHLLEW-----SW 1474

QY 597 -----ERIKDLPLQCYNNWTLTLGNK--WKTFETVHYLRSPKSGNPNFSTYEP 648
DQ 1574 GHNPRIERASMAQSMK-----IVIVQKTFW-----PQSLIIDY 1607

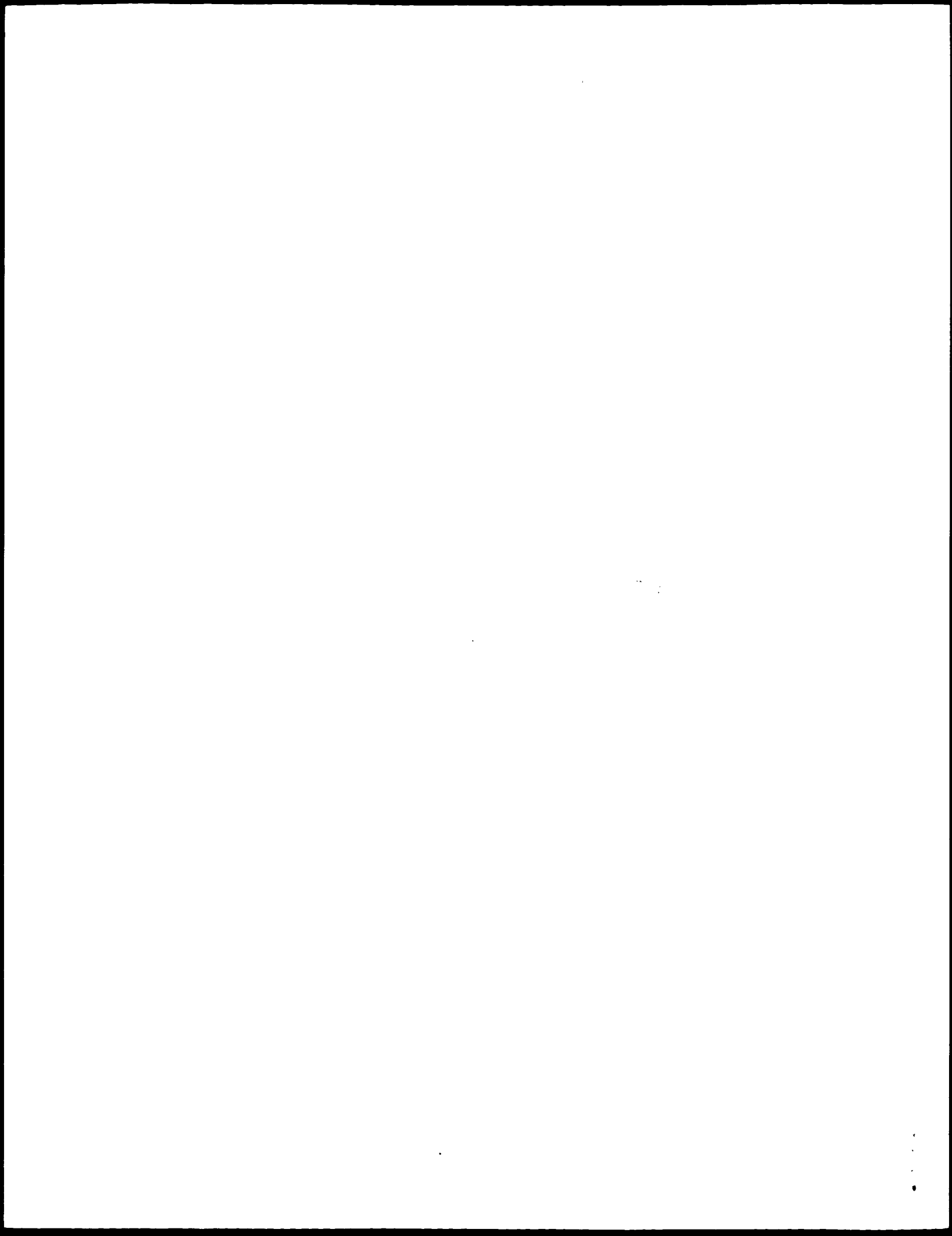
QY 649 LEFIDPSNLYAMK--INNIOVEGYSMHFDPEAL--POLLIGLIVPYT 692
DQ 1608 -----PNRLLYEMDSYLDYNDYNGYHIRKOVIASDLIIRIPALY 1649

```

Db 1356 ENGGTHREDFEFCAGCQICPQFLLANASKTCEDHDFCHLCS/SQHCYNMK CSERC 1413
QY 461 GTGTCGML-SQG-LCKPEVAEST-----DHVIGHE 489
Db 1414 -SCDTGMYLESQPTCKVTASESLLLLVASQKFIADSVTSQVNIYSLVENSYIVAVD 1472
QY 490 TD-----IQDLEMKYLLQKTDRRLEFVHAIFISNMRINSWFOPS--WKKRMILIT 536
Db 1473 FDSISGPIFWSEATQCKTWSAFQNGTUPAVVEUSSILITETIAIDWVGHNLYWTVYVALET 1532
QY 537 LKSNKYSSLVHMLGLSLQICLTKNSTLEPLVAVVVPFGSSSESSESWPMPVNSFFDW 596
Db 1533 TEVSK-----IDGSHRTVLISKNLINPGLA--LDRHMRNEFLPW --SUV 1573
QY 597 -----EPTKIDLPQCYNMWTLTLGNK--WKTFFFTVHLYLPSFKSNQPNENESIYVEP 648
Db 1574 GHHPRIERASMDGSMR-----TVIVQDKIEW-----PGLTIIDY --1607
QY 649 LEFIDSRNLGYMK--INNIOVFGYSMHFEDPAI-KRLLIQLIYNST 692
Db 1608 -----PHRELYENDSYLDYMDGCDYNGHHHPROVIAADLIIRHFAIT 1649

Search completed May 19, 2003, 10:35:16
Job time : 42 secs

100 US/00652477
From: Grotan
To: Grotan
Lats
QY: Grotan R.
Clarence C.
Edward M.
Grotan
No Human Calcium Sens. Protech. Fragments
No. thereof and DNA Encoding Same
EST: 106
DRESS:
Dr. Eugene Rorer Inc.
Sula Rd., 474
Cile
FBI
ONLY disk
close
W. System 7.5.1
6.1 (PatentIn)
N DATA:
SER: US/08/652,477
447
DATA:
SER: PT/US/65/152,4
N.V. 1995
DATA:
SER: US/08/444,840
N.V. 1994
DATA:
SER: US/08/487,414
JUNE 1995
FORMATION:
MATE: 29,699
NUMBER: A1856E US
INFORMATION:
454,9816
14,608
ID No: 84:
QUESTIONS:
509,948
133
13
MATE:
10000
US/08/652,477
Score 300,000,000 Length 4655;
487 Pred No. 100
487 50; Mismatch 100; Indels 187; Gaps 31;
CHGKTHREDFEFCAGCQICPQFLLANASKTCEDHDFCHLCS/SQHCYNMK-----YNSDFPS 320
AGTCKPOMCHSDFCQFCHLCS/SQHCYNMK-----YNSDFPS 1261
NYEL NESTIMILWTMSNEPATEVENSMSKOLELEKAKIVHKLFS 378
LYPHQNTNTHRAWLLW-----S 1300
PRCRLSYW-----ONENGLIG 417
KPSWQWUPLHNTWNLISVTHNHPNENDESPLNGNSGD 1355
QY: NUQVWTA-PIV-QASA-CLICAPINRRC 460




```

Db 181 AASVFIDRSTLPLLIHIIQIASTAKVTETPTGSLGSGSYENLEISVSSVVLGSPENKIQI 240
QY 241 QCLQVLIPIYLGPRVQAAI SV IACNSIGIPIKRIKRWGTHGCKKPIKPCNSMFDIAMP 300
Db 241 QCLQVLIPIYLGPRVQAAI SV IACNSIGIPIKRIKRWGTHGCKKPIKPCNSMFDIAMP 300
QY 301 ENLLRITETWKAYNSDFEESDEKFLMKRFLPMNYFLNTSTIMHLKWTMDSNFORRYEGLN 360
Db 301 ENLLRITETWKAYNSDFEESDEKFLMKRFLPMNYFLNTSTIMHLKWTMDSNFORRYEGLN 360
QY 361 SMKQLEKAKIVHKLFSLSKRCHKQPLISIPROPTSYWITPTIOSHYCNPNGLIGSPS 420
Db 361 SMKQLEKAKIVHKLFSLSKRCHKQPLISIPROPTSYWITPTIOSHYCNPNGLIGSPS 420
QY 421 EPTHSTCTNDQVCHVAFIPIVCHVQDASACLTICAPINRTRCGICNTGYMISQGLCKPVAE 480
Db 421 EPTHSTCTNDQVCHVAFIPIVCHVQDASACLTICAPINRTRCGICNTGYMISQGLCKPVAE 480
QY 481 STDHYIGFTDQLQLEMKYLLQKTDRIEVRHAIFISNDMRLNSWFDPSWKRKMLLTILKSN 540
Db 481 STDHYIGFTDQLQLEMKYLLQKTDRIEVRHAIFISNDMRLNSWFDPSWKRKMLLTILKSN 540
QY 541 KYKSSIVHMLIGLSIQICLKNSLTLEPVLYVYVNPFGSSHSSESWMPVNSFPDWERTK 600
Db 541 KYKSSIVHMLIGLSIQICLKNSLTLEPVLYVYVNPFGSSHSSESWMPVNSFPDWERTK 600
QY 601 LKAPLCQYNWLTIGKMKAFEEFEEVATKSKKSKSGHNGNESIYYEGRPEFIDE-SRNIGY 660
Db 601 LKAPLCQYNWLTIGKMKAFEEFEEVATKSKKSKSGHNGNESIYYEGRPEFIDE-SRNIGY 660
QY 661 MKTNNTOVQCYSMITDPEAIPDLIIQIDYPTQSGDSALLQILFTDPDVKLSPPQRRP 720
Db 661 MKTNNTOVQCYSMITDPEAIPDLIIQIDYPTQSGDSALLQILFTDPDVKLSPPQRRP 720
QY 721 LDLFSCLLRHRLKLTSEVVRIOALQAFNAKLPNTMDYDTTKLCS 766
Db 721 LDLFSCLLRHRLKLTSEVVRIOALQAFNAKLPNTMDYDTTKLCS 766

RESULT 2
US-09-864-761-43251
: Sequence 43251, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/297,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 2426336
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30

```

```

: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 43251
: LENGTH: 378
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL022143.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
: OTHER INFORMATION: EST_HUMAN HIT: AW955819.1, EVALUE 0.00e+00
: OTHER INFORMATION: SWISSPROT HIT: Q28661, EVALUE 9.00e-04
US-09-864-761-43251

Query Match 38.0%; Score 1550; DB 10; Length 478;
Best Local Similarity 74.7%; Fred. No. 9.6e+27;
Matches 278; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 395 RSTYWTUPTIOSFYCNENGLIGSPHSTHSCPTNPOVVCTAFLPCTVGVASACLTICAP 454
Db 395 RSTYWTUPTIOSFYCNENGLIGSPHSTHSCPTNPOVVCTAFLPCTVGVASACLTICAP 454
QY 455 DNRTGCTGNTGYMLSGIKKPFVAESTDHYIGFTDQLQLEMKYLLQKTDRIEVRHAIF 514
Db 455 DNRTGCTGNTGYMLSGIKKPFVAESTDHYIGFTDQLQLEMKYLLQKTDRIEVRHAIF 514
QY 67 DNSTRGCGNCPYVLAQGLCHPFAFSIENFGLFIDQGLKLYLQKQUSRIEVSIF 126
Db 67 DNSTRGCGNCPYVLAQGLCHPFAFSIENFGLFIDQGLKLYLQKQUSRIEVSIF 126
QY 515 ISNDMRLNSWFDPSWKRKMLLTILKSNKYKSSIVHMLIGLSIQICLKNSLTLEPVLYVYV 574
Db 515 ISNDMRLNSWFDPSWKRKMLLTILKSNKYKSSIVHMLIGLSIQICLKNSLTLEPVLYVYV 574
QY 127 ISNDMRLNSWFDPSWKRKMLLTILKSNKYKSSIVHMLIGLSIQICLKNSLTLEPVLYVYV 186
Db 127 ISNDMRLNSWFDPSWKRKMLLTILKSNKYKSSIVHMLIGLSIQICLKNSLTLEPVLYVYV 186
QY 575 PFGSSHSSESWMPVNSFPDWERTKLIRPLIKLPIQCYNWNLTIGLNAKLPNTMDYDTTKLCS 634
Db 575 PFGSSHSSESWMPVNSFPDWERTKLIRPLIKLPIQCYNWNLTIGLNAKLPNTMDYDTTKLCS 634
QY 187 PFGSSHSSESWMPVNSFPDWERTKLIRPLIKLPIQCYNWNLTIGLNAKLPNTMDYDTTKLCS 246
Db 187 PFGSSHSSESWMPVNSFPDWERTKLIRPLIKLPIQCYNWNLTIGLNAKLPNTMDYDTTKLCS 246
QY 636 SNTPNFTSYVEPLEFTDPSNICYMFINNLYVYVSMHFTFEAIEDLLIQILYPTQ 694
Db 636 SNTPNFTSYVEPLEFTDPSNICYMFINNLYVYVSMHFTFEAIEDLLIQILYPTQ 694
QY 247 SCSALLQILIRDRVKNLSPPQRRFLCLLSCLLRHRLKLTSEVVRIOALQAFNAKLP 306
Db 247 SCSALLQILIRDRVKNLSPPQRRFLCLLSCLLRHRLKLTSEVVRIOALQAFNAKLP 306
QY 695 SCSALLQILIRDRVKNLSPPQRRFLCLLSCLLRHRLKLTSEVVRIOALQAFNAKLP 754
Db 695 SCSALLQILIRDRVKNLSPPQRRFLCLLSCLLRHRLKLTSEVVRIOALQAFNAKLP 754
QY 755 NTMDYDTTKLCS 766
Db 755 NTMDYDTTKLCS 766
QY 766 NTMDYDTTKLCS 766
Db 766 NTMDYDTTKLCS 766

RESULT 3
US-09-864-761-43252
: Sequence 43252, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/297,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 2426336
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30

```



```

: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 0.99
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL - 0.99
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 1.3
: OTHER INFORMATION: EST_HUMAN HIT: R15400 1, EVALUOE 5.00e-37
: OTHER INFORMATION: SWISSPROT HIT: P16581 EVALUOE 6.60e-01
US-09-864-761-38031

Query Match
Best Local Similarity 8.0%; Score 127; DB 10; Length 78;
Matches 55; Conservative 10; Mismatches 13; Indels 3; Gaps 0;

QY 243 IQVLLPYLGERFVQAALSYIAINSEGEFTIKENQDWTHGPKPEKPCNCPMDICAMEEN 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 IQVLLPYLGERFVQAALSYIAINSEGEFTIKENQDWTHGPKPEKPCNCPMDICAMEEN 60

QY 303 LIRITETWKAYNSDPRES 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 LLIQDSWATHNRQFEES 78

RESULT 5
US-09-864-761-38134
: Sequence 38134, Application US/09864761
: Patent No. US2002/0048763A1
: GENERAL INFORMATION:
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED, SINGLE EXON NUCLEOTIDE A'ID PROBES USEFUL FOR
: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263,6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/226,459
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-05-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-30
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

```

```

: SEQ ID NO 38134
: LENGTH: 69
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL022143.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 0.99
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 1.2
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 1.4
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL - 1.5
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1
: OTHER INFORMATION: EST_HUMAN HIT: U72043.1, EVALUOE 5.00e-16
US-09-864-761-38134

Query Match
Best Local Similarity 5.4%; Score 219; DB 10; Length 69;
Matches 46; Conservative 10; Mismatches 7; Indels 6; Gaps 1;

QY 144 EESLIIFVDKPKLKPAGSGSSTI-----NSSSVTLTLHQAASYFIHQUSTIRLHH 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 EESLIIFVDKPKLKPAGSGSSTI-----NSSSVTLTLHQAASYFIHQUSTIRLHH 60

QY 198 IOIATAIK 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 IOIATGAIK 69

RESULT 6
US-09-991-496-121
: Sequence 121, Application US/09991496
: Patent No. US2002/0149285A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Campos-Neto, Antonio
: APPLICANT: Webb, John R.
: APPLICANT: Dillon, David C.
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Bhatia, Ajay
: APPLICANT: Coler, Rhea
: APPLICANT: Probst, Peter
: APPLICANT: Haddon, Mark
: TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
: TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
: FILE REFERENCE: 210121.420C9
: CURRENT APPLICATION NUMBER: US/09/991,496
: CURRENT FILING DATE: 2001-11-20
: NUMBER OF SEQ ID NOS: 137
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 121
: LENGTH: 709
: TYPE: PRT
: ORGANISM: Leishmania major and chagasi
US-09-991-496-121

Query Match
Best Local Similarity 3.0%; Score 123; DB 9; Length 709;
Matches 75; Conservative 26; Mismatches 117; Indels 126; Gaps 16;

QY 203 TAIKVTETRTGPGGCS-- . . . . . KVNFSVSVSVLVGSEPHKPKLGLGLGV 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 354 TAMKUPYINVAVLGGASCTTCFNETACLETFSEYML-----PIMVSLTGLQC 404

QY 246 LLEPY-----LQRFVQAALSYIAINSEGEFTIKENQDWTHGPKPEKPCNCPMDICAMEEN 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 404 LQRFVQAALSYIAINSEGEFTIKENQDWTHGPKPEKPCNCPMDICAMEEN 464

QY 278 CWCHGCPKEFEC NCFS-- . . . . . MD LGAMEENLRLTETWKAYNSDPREE 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 464 CWC-- . . . . . HLPKSCFSAAGCTCTGCANCGLVGACVACGCFNFCSC----- 507

QY 300 SHPKPKPEKPCNCPMDICAMEEN-----TETWKAYNSDPREEFVQUSTIRLHH 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 508 SDANK--CTGAPNYLTLPLLTGSPVACNIEHMQDQPIKQGVSP----- 555

```

```

APPLICANT: Chen, ping w.
TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
FILE REFERENCE: UAB-17403/22
CURRENT APPLICATION NUMBER: US/10/047,676A
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/627,376
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 990
TYPE: PRT
ORGANISM: Streptococcus mutans
US-10-047-676A-7

Query Match      2.8%  Score 115.5;  DB 12;  Length 990;
Best Local Similarity 18.9%  Pred No. 0.55;
Matches 154;  Conservative 125;  Mismatches 270;  Indels 265;  Gaps 42;

Oy 119 TLQOITENIKKYC-----THP-ILSALIGGFFSLTFIVKPKISKPAKGSUHS 169
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 139 TLKVANNALGKSNDEWLLDTSHFGLMNSPSCIPEDITV-----PSNQLIDVING 190
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 170 SS--VTLETLHQAASYFIDPDSTLPPHHLQIATAIK-----VIEPICHACCSYEN- 222
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 191 TEEDISYQILIDIDIAEKFSOSSDIK-----EYIQLIKKEFLITEIKFSLID-----DNP 241
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 224 TSVSSVSLVSPENKILGK--GVLEFVLEPFVAAALSYAAGNSELKERN 276
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 242 LQWFINILERDONSLELEKLEIKAMLOYDTDRNIGGNNSI ALENKMSQIVKANAYL 301-
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 277 --DCWCHGKPPPCNCPMDIOAMENILITETW-KAYNSDDESDPKIPMKR----- 329
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 302 RVDLYHAEFLAQHTKSSL-----QNLKVLSSSSAVNSCKEIKNVHEKFIARYGYE 355
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 330 --LPMNYFLNTSTIMHLWTMDSNFORBYEQLE-NSMKOLFLKAKIVHKL---PSISKRC 383
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 356 QLVFQULLINS-----TSGLGPPGYSQTFVSKANNELSPKRTIEFQPPFKAIRD 408
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 384 HKQELISLFRQRTSY WLRIQSSELYEN-----GLLASESEELHS-IC 428
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 409 GKELISLDDIKDLNFDTBOUIGELNCHYNFASKNLEVSIGVSOMGNTGPHHS KL 467
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 429 ENLAVVIAIPQIVG-ASALICAKNNHFGGNGY-----MSSGIAAPVAAS 481
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 468 FN-----IVTKNVKTKETFEAVENTILQNEVEFEGRGNNIMISNL KSHQLLE 520
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 482 IDWYDQFQDQD-----FMKYIQIDQFPIVHAFIENKMPNENWEPQWPKRM 534
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 531 KNYTKKEMSINDIYVRAISSELYFYSKYKAV-----IVVMNM--FNYINSKILKEL 574
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 535 LTKSNKYKSSLVIMILGL-----SLQICLTK 561
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 575 LEVNSDQF-NITPITIGSISDYNHVAIYKDIITKPELWNERKSEAKILASIKMWIT- 612
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 562 NSTLEPVL-----AVYVNPFGGSHSEFMPVENSE-----POWERIKIDI 603
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 644 NNNVPPFVPMKHTPILVLLSFTILLIMLEFKKHSFLLLELVHSVNTNTEILELVV 692
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 604 PL-----QCYAWTITL-----CNKWKTFETVHI-----YLR- 630
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 693 PFKSDVNAHQIYHYACNIVILEISGSKKAYFYAKIVYNNKQD-SELOKYEPLIAYIKL 752
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 631 -----SRKSNPGNE-----SIYEPELEFIDPSNGLYMKINNITGVFG 670
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 753 PENIQWFYIKDDGKDSIRIRTRYVEDKQIVQIYSPELEWAIKAR-----KNITQSG 805
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 671 YSM-HFDPFAIR-----DILQLDYPYTGGSQDSALQLLEIRDR 709
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 806 YEISEYIPESARYGKKYSSIIHSFFYYDSILLDLLO-----KKAETIEVRTS 854
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 710 VN-----KLSPPGQRRL--DILFSLQRHRUK 744
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

HKQELISLFRQRTSYWLRIQSSELYENGLIGSFSHETHSCIPND 431
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
KLSLWSPENKILGKSNDEWLLDTSHFGLMNSPSCIPEDITV 608
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
VLASATLICAPUNIRIRGIONKIMLSGRLCK 475
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
EHNICVNGDSIRIRTRYVEDKQIVQIYSPELEWAIKAR 642
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Title: US/08/74924
(1120A1)

Steven G.
Nelson, Antonio
John R.
David C.
Yasir A.W.
Ajay
Rhea
Peter
Mark
LEISHMANIA ANTIGENS FOR USE IN THE
THERAPY AND DIAGNOSIS OF LEISHMANIASIS
121.42068
NUMBER: US-08-07;
2001-06-04
3:122
for Windows Version 4.0

is a major and charasi

3.8%  Score 123.18 10;  Length 709;
21.8%  Pred No. 0.07;
Conservative 24;  Mismatches 117;  Indels 126;  Gaps 16;

GAPS -----NUNMUNVAVSVQSPENKIQIGQLQV 245
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
LQCSAGTIGENFACIETIRQYLM-----LIMCSLIGLQV 403
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
-----LQEFVQALYIANF EREFTK-----END 277
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QTLNNDQYGLISSVVRGSAVKEFTEVYANVKVCTGCEEPINNM 453
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
NOPS -----MPP-LOAMERLEKIETTWKAYNSDDEP 319
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ASCPDAITCTQANGYGLVAGAVRQEPNCS-----D 507
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
MNYEN-----ISTIMHWMISNFORRYQIENSMKOLFLKAK 371
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QNYVLELLIGSPVACNIEHQMUTQPTSPKQCVSP-----555
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
EPFLIEEPFPTIYWTSFELLYMEHILASTFEETISATVND 431
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
KLSLWSPENKILGKSNDEWLLDTSHFGLMNSPSCIPEDITV 608
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
VLASATLICAPUNIRIRGIONKIMLSGRLCK 475
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
EHNICVNGDSIRIRTRYVEDKQIVQIYSPELEWAIKAR 642
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

om 08/10047676A
4105A1

Title:
Page W.

```

Db 855 LSIIRFMMKLSLQDQKLIKLNFDG--KHKLK 886

RESULT 9

US-09-991-496-22

; Sequence 22, Application US/09991496

; Patent No. US20020169285A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-11-20

; NUMBER OF SEQ ID NOS: 137

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 23

; LENGTH: 320

; TYPE: PRT

; ORGANISM: Leishmania major

US-09-991-496-22

Query Match

Best Local Similarity 32.8%; Score 114.5; DB 9; Length 320;

Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;

QY 383 CHKQPLISLPROPTSTYWLTRIQSELYNENGLLSFSEETHSTCPNDQVVCIAFLPCT 442

Db 176 CRISDACSVPNCKKCFETSTPL-----CAEIDGYSLALATSCSSPTIQ-----PCE 223

QY 443 VGDASACILTCAPNRTCTCTGTCNTGYMISGLICK 475

Db 224 V---EHCNTCVNGDSTRCAVCNTGYVYVSDGCK 253

RESULT 10

US-09-991-496-55

; Sequence 55, Application US/09991496

; Patent No. US20020169285A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-11-20

; NUMBER OF SEQ ID NOS: 137

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 55

; LENGTH: 320

; TYPE: PRT

; ORGANISM: Leishmania major

US-09-991-496-55

Query Match

Best Local Similarity 32.8%; Score 114.5; DB 9; Length 320;

Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;

QY 383 CHKQPLISLPROPTSTYWLTRIQSELYNENGLLSFSEETHSTCPNDQVVCIAFLPCT 442

Db 176 CRISDACSVPNCKKCFETSTPL-----CAEIDGYSLALATSCSSPTIQ-----PCE 223

QY 443 VGDASACILTCAPNRTCTCTGTCNTGYMISGLICK 475

Db 224 V---EHCNTCVNGDSTRCAVCNTGYVYVSDGCK 253

RESULT 11

US-09-874-923-22

; Sequence 22, Application US/09874923

; Patent No. US20020081320A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-06-04

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 22

; LENGTH: 320

; TYPE: PRT

; ORGANISM: Leishmania major

US-09-874-923-22

Query Match

Best Local Similarity 32.3%; Score 114.5; DB 10; Length 320;

Matches 30; Conservative 8; Mismatches 40; Indels 15; Gaps 3;

QY 383 CHKQPLISLPROPTSTYWLTRIQSELYNENGLLSFSEETHSTCPNDQVVCIAFLPCT 442

Db 176 CRISDACSVPNCKKCFETSTPL-----CAEIDGYSLALATSCSSPTIQ-----PCE 223

QY 443 VGDASACILTCAPNRTCTCTGTCNTGYMISGLICK 475

Db 224 V---EHCNTCVNGDSTRCAVCNTGYVYVSDGCK 253

RESULT 12

US-09-874-923-55

; Sequence 55, Application US/09874923

; Patent No. US20020081320A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Webb, John R.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea

; APPLICANT: Probst, Peter

; APPLICANT: Brannon, Mark

; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE

; FILE REFERENCE: 210121.42008

; CURRENT FILING DATE: 2001-06-04

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 55

; LENGTH: 320


```
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/245,041
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 60/093,630
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: 60/104,978
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-238-17

Query Match      2.5%; Score 100.5; DB 10; length 1350;
Best Local Similarity 20.68; Pctd. No. 18;
Matches 70; Conservative 34; Mismatches 76; Indels 159; Gaps 20;

QY 253 ERFVQAALSYIA'NSRGE-FTG--KENDG-WP--HGGPKPEEC-----NCP--- 292
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 690 ECFSKRTLDHRCDOHTDCYSCYCTANTNDCHWCNDHCVPRNHSCEGQISIFRYVNCPKDN 749
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 293 -----SMDIQAMEENL-----LRITETWKAYNS 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 750 PMYYCNKKTSCRS'ALDQNGWEPENQETALPENITG'JHMLV'NSCLKITTAKENY-- 807
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 DFESEDFKLFMKRLPMNYFLNTSTIMHLWTMDSNFORRYHQLNSMKQLFLKAOKIVHK 375
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 808 -----DNAKLFERN-----HALLASLITQ-----KKVEFVLKQL-----RIMQS 842
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 376 LFSLSKRCHKQPLSLPQRTSTYW--LTRIQSELYCNEH-----G 414
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 843 SOSMSK-----LTIITWVGLRKINVSYWCEWEDMSPTNSLLQWMPSEPSDAG 889
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 415 LLGSESE-----ETHSCTCPNDQVVC-----TAFLECTVGDASACL 450
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 890 FCGILSEPSTRG'KAATCINP'INGSVCEPPANHSKQCP'PCALPTACGDTSG--SSECM 948
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 451 TCAPIDNRTHC-----GTCNTGYMISOGICKPHE 477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 949 WGS--NMKQCVDSNAYVASFPFGQCMWYIMS--TCPPE 983
```

Search completed: May 19, 2003, 10:35:44
Job time : 34 secs

bioinformatics version 1.1.4 (p.4578)
Copyright (c) 1993-2003 Compuser Ltd

search using sw model

2003.10.28.53 : Search time 47 Seconds
(without alignments)
[966,787 Million cell updates/sec]

seq: 962,129,45

SRAGAELESLMALWEW : AFNAKLNTMDYDTKLCS 766

seq: 2,000,000

seqs: 96144222 residues

displaying chosen parameters: 283224

2003.10.28.53

on March 03

on March 100%

for first 45 summaries

100%

100%

100%

100%

number of results predicted by chance to have a
in or equal to the score of the result being printed,
analysis of the total score distribution.

SUMMARIES

ID	ID	Description
1	109052	hypothetical prote
2	852835	hypothetical prote
3	2047	large tegument pro
4	144178	complement C7 prec
5	843	large tegument pro
6	2077	GTPase-activating
7	114	botulinum neurotox
8	86774	signal recognition
9	1105	hypothetical prote
10	140644	interleukin-6 sign
11	599	ORF MSV240 leucine
12	847928	hypothetical prote
13	2085	large tegument pro
14	149755	probable membrane
15	44257	perforin 1 precurs
16	527	1pr homolog - 1pr
17	20270	hypothetical prote
18	44	probable RAS GTPas
19	254	integrin, band 3 p
20	141933	conserved hypothet
21	845743	genome polypeptide
22	437183	hypothetical prote
23	11820	exodeoxyribonuclea
24	1522	cytolysin precurs
25	145842	dynamin heavy chain
26	43	isolucine-tRNA li
27	14083	keratin Ha3-11. ty
28	147459	hypothetical prote
29	2962708	probable membrane
30	867202	

ALIGNMENTS

RESULT 1

109052

hypothetical protein DBCCRL - human

C:Species: Homo sapiens (man)

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #tax_change 21-Jul-2000

C:Accession: T09052

R:Habuchi, I.; Luscombe, M.; Elder, P.A.; Knowles, M.A.

Genomics 48, 277-288, 1998

A:Title: Structure and methylation-based silencing of a gene (DBCCRL) within a candid

A:Reference number: 216537; NCID:98207242; PMID:9545642

A:Accession: T09052

A:Status: preliminary; translated from CH/EMBL/DBHJ

A:Molecule type: mRNA

A:Residues: 1-761 <HAR>

A:Cross-references: EMBL:AF027734; NID:q3041876; PID:AA049691.1; PID:q3041877

A:Experimental source: tissue type: brain; developmental stage: fetal and infant

C:Genetics:

A:Gene: DBCCRL

A:Map position: 9q32-q33

A:Note: H3089A

C:Superfamily: human hypothetical protein DBCCRL

Query Match 53.0% Score 2161; DB 2; Length 761;
Best local Similarity 52.3%; Pred. No. 1,16-148;
Matches 406; Conservative 134; Mismatches 209; Indels 28; Gaps 10;

QY	1	MWRSRAGAELESLMALWEWIALS-LHCWLVAAVAVSQQHATSI-FWLLISLQKGFHPISQE	59
Db	1	MNRRE---VELLYFLWGRINQVPSH---DEPAGTIDHVSKEFWLLISLQKGFHPISKS	53
QY	60	YTFVVDHSQGGSTPKYTFEFGRWKVNIAVERNNFAGSPFLAEPFNNILIGKPT	119
Db	54	YLFSEVHEHFGFTYKYTFEFAEKVRNTATERRDLVHPVLMHEFQSRILDRRPT	113
QY	120	LCQITENLKFYTHPELLSAILGTFEFLTFVPRPFI-KFAEISISTNSSSVILELHQ	179
Db	114	TQFPIPLKKGTHLHLSAILGGLALLIMYMKSM-----DKNSNATQSVFALHQ	166
QY	180	LAASYF-IDRDLRLRHIIQIATAIKVTFTRTPISNYNIISVSVSVLSQSPENIQ	249
Db	167	LASSYFVDEGIMPRFHEITQISTGAIKVTETGTGLYNSVNIISVSVLSQSPENLR	226
QY	240	LCQVAVLHLYQERVQALSYLAQNSGDF-LCAENQWCHQCKPFGNPTSMIDQAM	299
Db	227	LQSLQILFFYLQERVQSALSYIMNGEGEYDQNSQGRQVAFEPQVNPFTIDQIM	286
QY	300	EENLLR-TETIKAYNSDFEESDEFKFLMKRLPMVFLNTISTIMLWTMGSNFORRYOLE	359
Db	287	EYTLANMAKSWALAYKLENSDFKSFPMKRLSNHPLIGSLDHWGNWLOLQNKYKLO	446
QY	360	NSMKQLFLKAKIVHKLFLSLSKPCHQVPLISLPQAFISTYWLTF-IQSTLYNENGLASF	419

hypothetical prote
hypothetical prote
complement C7 prec
hypothetical prote
protein C44B4.1a 1
perforin precursor
hypothetical prote
240K tegument prot
hypothetical prote
cytochrome b450 -
resilin precursor -
serine/threonine-s
receptor-like prot
kinesin-related pr
probable RNA repl

Db 347 SATPAORQKQRTARRKLPGLSVKCRHNPQHQPRTTQQMLARVOSLLIYCNCFWCTP 406
 QY 420 SEETHSITLQFNQAVVTAFLPCTVIGASALITAFQNPTRGCTNTGMYMLSGGLCKKEVA 479
 Db 407 LESQSCVWJGHTLLQRPPIVIGNNSTCTMSLANISLQSCNKGKYLGRCEPNV 466
 QY 480 ES--TDHYIGFDTL--QDLEMKYLLQKTRDRFEVHAIFISNDKRLNSWFOFSPWRKRLG 535
 Db 467 DSRSEQFISFETDLPQDLEKLYLKQKMSRLYVHTTIFISNEIPDTEFDPWRKMSL 526
 QY 536 TLKSNKYKSSLVMLILGLSLQILCKNSTLEPVLAVYVNPFGGSSSESFWPVPVNSEFPD 595
 Db 527 TLKSNKMDPIMVTCMSMRCKMNSLDPFMYVNPFGSGHSEGNMPFQFQYPR 586
 QY 596 WERTKLDLPQCYNWTLTGNKWKTFEFVHIYLSRSKNSGPNGNESFYEPLEFIDPS 655
 Db 587 WEKIRLQ--NSQCYNWTLLGNKWKTFEFVHIYLSRSKNSGPNGNESFYEPLEFIDPS 644
 QY 656 RNLYGKNNKINNOVFGYSMHFDPHAIKDLILQDYPYTGQSQ-----DSALLQILKIKRVN 711
 Db 645 KPGFYKISVAVVQVGYSLDFENALLPSAVQVYNSYTGQFQFYSVSSVWMLLLLTIPCRN 704
 QY 712 KLSF--EFGPRIDLPSTLPRILKLTSEVVRVQSLQAENAKLPNTMDYDTIKLG 765
 Db 705 KIAPPVAKPKPQDLPSCLKRLKLTNSSEITRVNHALDLYNTEILKOSDOMTAKLC 761

RESULT 2

S52835
 hypothetical protein, YMK075w, Y-mast (Saccharomyces cerevisiae)
 N: Alternate names: hypothetical protein YMK075w, Y-mast (Saccharomyces cerevisiae)
 C: Species: Saccharomyces cerevisiae
 C: Date: 14-May-1995 #sequence_revision: 61-Sep-1995 #text_change: 1-Apr-2002
 C: Accession: S52835
 R: Pearson, D.; Bowman, S.
 submitted to the EMBL Data Library, April 1995
 A: Reference number: S52814
 A: Accession: S52835
 A: Molecule type: DNA
 A: Status: preliminary, translated from GS/EMBL/DBJ
 A: Cross-references: EMBL:Z48452; NID:g763008; PIDN:CA88800.1; PID:g763022; GSPDB:GN0001
 A: Experimental source: strain AB972
 C: Genetics:
 A: Gene: MFS:YMK075w
 A: Cross-references: SGD:S0004680
 A: Map position: 13R

Query Match 3.38; Score 135.5; DB 2; Length 684;
 Best local similarity 18.68; Pred. No. 0.1;
 Matches 128; Conservative 86; Mismatches 221; Indels 255; Gaps 36;

QY 76 KIYPERGRKVVNLAVERPNTLGSPLAP-----EFPNIRLLGPRPTLQITENLIK 129
 Db 122 KIKRE-SLMYRKNIQCQNN-----SMHTLVTSKRPTQVTKNPQDNRNDIKTFITENMTE 177
 QY 130 KYGTHPLLSATIGGCKSLTIPVDKKKLSKRAGSDSTTSSSVTLETILHQLAASYFIDRD 189
 Db 178 ESN-----IPSTICWNQDIT-----NTRDPEPESDPDNKKLSNIRTKIILSNATY----D 225
 QY 190 STLRKLIHQIATAIKVTEIKTIGPLQCSNYUNLUNSVSVQSPENKIQIQQIQLIPI 249
 Db 226 SKSLKPCQNSIKSTS-----NASKKIPROKNN-----TID 256
 QY 250 YLQKRFQALSYLACNSBUEELKENDQWJGKGFKECNC--PSMDLQAEENLRIT 307
 Db 257 FNRHDFCS-----ACNQSQSPIC-----CUTCPKSPHFICLDPPIDPNKIPKQDMHCN 304
 QY 308 ETWKAYNSDFERSDEPKLPMKRLPMNYFIINTSTIMHLTWDSNFORRYEQLEENMKQLPL 367
 Db 305 EC-----KPKIFI-----NNSMATLKIESNFIK-----QNNNVKIFA 337
 QY 368 KAKQIVHKLISLSKCKCHKQPLISLPHQRTISIVYLTIRIOSPIYCNNGILGFSSETHSCT 427

Db 338 KL-----LPNIDSHNPKQ--FQIDPNYIKETPPAVK-----TQSRGOYSDF----- 375
 QY 428 QNDQVAVVCTAFLEPCVIGASALITAFQNPTRGCTNTGMYMLSGGLCKKEVAESTDHYIG 487
 Db 376 --NDKRIPLT-----LFTSY--QGSITK-----LDSY-N 402
 QY 488 PFTDQLQLEPMKYILQKTRDRFEVHAIFISNDKRLNSWFOFSPWRKRLITKSNFYKSSLV 547
 Db 403 PDTHIDNSGKELI-----CYKNOTRLGSSSIPE--NSRLIMT----- 439
 QY 548 HMTLSLSLQIGCTKKNSTLEPVLAVYVNPFGGSSSESFWPVPVNSEFPDWEPTKLDLPLQ 606
 Db 440 -----CUCQCT-----PWHLDQVPRASFKN----- 459
 QY 607 CYNWTLLTGNKWK-----TFEETVH-----IYLPSPKSKSGPNGNESFYEP 648
 Db 460 -----LASKKKOPLHSPTKYKKIHHQEDNSVNYVYVWKQQL-----INFRKSLIYEP 508
 QY 649 LEFIDPSRNLGYMKINNTQVFGYSMHFDPHAIKDLILQDYPYTGQSQSALLQILKIRD 708
 Db 509 LQ-----KIGYCNNGNQLIETTSITDQD---PQDCKKITQIDENS:KYDEF---D 553
 QY 709 RVNKLSPPGQPLLELFSCLLRHRLKLTSE 738
 Db 554 KIYKSMWOKRKLPQFQESLIDKIVNSGQ 583

RESULT 3

T44178
 large tegument protein U31 [similarity] human herpesvirus 6 (strain 229)
 C: Species: human herpesvirus 6
 A: Variety: strain 229
 C: Date: 21-Jan-2003 #sequence_revision: 21-Jan-2003 #text_change: 05-May-2000
 C: Accession: T44178
 R: Dominguez, G.; Dominguez, T. F.; Stamey, F. P.; DeWurst, S.; Inoué, N.; Bellett, P. E.
 J. Virol. 74, 8040-8054, 1999
 A: Title: Human herpesvirus 6B genome sequence, coding content and comparison with human
 A: Reference number: 222734; MIM:10482553
 A: Accession: T44178
 A: Status: preliminary, translated from GS/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-2077 <DOM>
 A: Cross-references: EMBL:AF157706; PIDN:AA049645.1
 A: Experimental source: strain 229; variant B
 C: Genetics:
 A: Note: 031
 C: Superfamily: varicella-zoster virus 240K tegument protein

Query Match 3.28; Score 129; DB 2; Length 2077;
 Best local similarity 21.08; Pred. No. 1.5;
 Matches 172; Conservative 109; Mismatches 248; Indels 292; Gaps 44;

QY 57 SQEYTDVPIVSPQSGSTPKYIYRFGPKVKNLA-----VEPRNFGSPPLAPAEFFRNI 111
 Db 584 SEKLNTIHTLNKKT---KIVQD-----NYNIAAGFMVITELNCFAYI---VNLYFNI 633
 QY 112 KILGR-----KPLQQTITENL-----IKKYGT-IF-----LLSAILGSESLTIVDRKKL 156
 Db 634 EVLKHGSLNTELDQVEKLYDNTQFLRFGTSHFNINLSNFTLSIRKKEVDFYNSQKP 693
 QY 157 SKRPG-----SDSTNSSSVTLETILHQLAASYFIDRDSTIRPHHIGIATAIKV 207
 Db 694 SDPASEIIAIESILADSPKNTIVNIEMI-----KSQLEELGKMEISTENKQ 742
 QY 208 TTRTGP-----LQCSNYDLDSVSS-----VLVQSPF-----NRKIQ 240
 Db 743 TAFITKQIHGQELPIYVDFLHLSAYNLPNTITVKNLHIFILEKRPDIAAIIHDKIQ 801
 QY 241 LQSLVLIPIYIGEPFV--LQALSYIA--NSFIEI--KENLAW--HGFKE--RNF--MFCAM 299
 Db 802 SLIDICIDMLNDIIVPEQTSTVL-----ELV-----DLFEN-----STETAL 841
 QY 300 EENLILITETWK--AYNSDFESDEPKLPMKRLPMNVPIINTSTIMHLTWDSNFORRYEQ 357

K:YAN:KILEEEDDLAJFIINS.....EQ 877
K:IVHKP:SLSKRCHKOR:LSL:R:LSYWL:RIOSFLYCNENGLIC 417
K:KPTLMCHIFFLYSAF:ITAEVNWIGR..... 916
NDQVV:TAFLPCTVDSASV:TPAPN:RFGTONTGYML:SG:QKPE 477
E:ATAP:KHALQ:---TCKPE 947
TOLGLENKYYL:KTPPE:IEVHA:ENTMLNSWFDSPWKRML 535
MEEQMKTAEN:KFKH:PIENTLEKRLND:-----LIL 987
.....SVVHMLG:LSIO:ILIKNSTLEPV 568
VILNL:LETEKULQDNML:ILH:THALL:VLKDIETTSIKTSILILEKI 1047
RSEWTFMEVNS:SPFWF:TF:G:G:YTWLILANKW KIEFETVH 526
KPECKYIPE:-----LSFLSKFKK:TFCT:--- 1082
NINSIYVELEP:DPKSLNAYWK:NN:LVNGYSMHEDPEAIRDLILQ 686
MOKSIIF:---LTKATSTN KYTP:ISH SVYRFLNLYEERTTELKKE 1134
ALQLEIPE:PVAK:SPH:G:R:G:SPLEPHRI:----- 732
AVAEKKILLSQDARTKLYLVNTHILKEIKNTOFFE 1184
VRIQSALQAFNAKLENTMC:Y:TEK: 764
ANKENLEAFI:NNHNAHI KAKM:---NE:IK: 1225
.....[untranslated].....human
.....sequence revision to Sep 1999 *text_change 31-Dec-2000
34072: C54072
stavaril, D.N.; Muller Eberhard, H.J.; Fey, G.H.
1988
of human complement component C7 and the C5b-7 complex.
744: M01D:86087145; F01D:1435508
104597: NID:q174715; E01R:AAA51861.1; PID:q179716
R: G:Scripto, R.G.
205: 1993
the human C6 gene.
3072: M01D:934291176; F01D:1435508
sequence not shown; not compared with conceptual translation
21: 87-22 143 (HRC)
X72140; NID:q142769; H: X72191; NID:q112800; CR: X72192; NID:q3128
ted from NCBI backbone (NID:IP:134085)
n is not annotated in GenBank entries HSC7X2, HSC7X3, and HSC7X4;
sequence not shown; not compared with conceptual translation
X72194; NID:q142769; H: X72194; NID:q142798
ted from NCBI backbone (NID:IP:134087)
n is not annotated in GenBank entries HSC7X13, and HSC7X14, relea
mers, M.; Hess, D.; Furmanek, A.; Miroschnichenko, O.
86-4274, 1999
nal components of the complement system are C-mannosylated on mul
4362: M01D:20020247; F01D:143551849

A:Note: identification and location of C-mannosylation sites by mass spectroscopy
C:Genetics:
A:Gene: CDH:C7
A:Cross-references: CD8:119046; OMIM:217070
A:Map position: 3p13-5p13
C:Complex: combines with complement C5b 6 complex to form the complement C5b-7 complex
C:Function:
A:Description: as a component of the complement C5b 6 complex, assists in the polymer
A:Pathway: complement pathway
C:Superfamily: complement c6; complement factor H repeat homology; KGF homology; LDL
C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein
F:122/Domain: signal sequence *status predicted *sig
F:23 843/Product: complement component C7 *status predicted *MAT
F:24 86/Domain: thrombospondin type 1 repeat homology *THR1
F:85-119/Domain: LDL receptor ligand-binding repeat homology *LDL
F:455-486/Domain: KGF homology *KGF
F:459-546/Domain: thrombospondin type 1 repeat homology *THR2
F:571-626/Domain: complement factor H repeat homology *FH01
F:631-688/Domain: complement factor H repeat homology *FH02
F:69/Modified site: 2' mannose 6-phosphate (M6P) *status experimental
F:202-754/Binding site: carbohydrate (Asn) (covalent) *status predicted
F:503 506/509/Modified site: 2' mannose 6-phosphate (M6P) (covalent) *status experimental
Query Match 3.18; Score 126; DB 1; Length 844;
Best Local Similarity 20.48; Pred. No. 0.67;
Matches 107; Conservative 57; Mismatches 153; Indels 298; Gaps 29;
QY 32 VAAVSQDHAISPPDWLLISUKGPFR-----SQHYDFDVRSGCESTRYKLYRECRMKV 86
DB 208 VKHIST-HTSS... SKPSPFFSSSSSSPSYTSHTNFTHKPSYQLLY-----V 253.
QY 87 NNJAVERNELGSP---LPLAEFFERNILGRRPTLQOIT--ENLKKYGHIFLLSATL 141
DB 254 EN-TVEAQPINNPPFLG:LAEPFKK---LSHLSLYDYSAIRKLDYGYGHYQSGL 309
QY 142 GGEESLTFYDKPKLSKPAEGSDSTTSSSVTLETLHLAASVYIDPDETEPLHHIQT 201
DB 310 GREYPV-FYVDSERIKQ-----NDENSEEKKFYSSWMHVVKFS-----H--- 351
QY 202 STAIKVFETPTGPGSNVINIDSVSSVIVGSPNKKIQIGLQVLPNYLQHPFVQ--- 257
DB 352 -----GCKELNALKAAS---GTQNNV-LRG-----EPFIRGGCA 382
QY 258 AALSYYAL-NSEFEICKENLWCHGPKFPAEN-PSMDICAMEENLLITETWKAY 313
DB 383 GFISGLSYLELDNPAEN--KRRYSAMAFSVINLP-----CVIKQKLLPIYE----- 426
QY 314 NSDFEESDEKFLPMKRLPMNYFLNTSTIMILWTMSNFORRYELLENMKOLFKAQIV 373
DB 427 -----LVKEVPC-----ASVKKLYLK--NAL 445
QY 374 HKLPSLSKRCCHKOPLISLPHQRTSTYMLTRIOSFLYCNENGLICSPKTHSCIPNPOV 433
DB 446 EYLDPEFDPCHCRP-----CQNGGL--ATVEGH-CLCH----- 476
QY 434 VCTAFLPCTVDSA-----SACLTCAPELNRCTGNTGYVLSQ 472
DB 477 -CK---PYTGAACPGCVLVGNAGCGVGGWSSWSPGVGCKRIPSPEDNNPPPSGCG 532
QY 473 -LCKPEVAESTHYIGFTLDQLEMKYLLQKTRDRIEVIHAFIS 516
DB 533 RSCVGE*TES-----TQCEDEELEHL-----RLLEPHCFPLS 564
RESULT 5
T43991
large tequment protein [imported] - human herpesvirus 6 (strain HST)
C:Species: human herpesvirus 6
A:Variety: strain HST
C:Date: 23-Jan-2000 *sequence_revision: 21-Sep-2000 ****_change: 21-Jul-2000
R:Accession: T43991
C:Sequencer: Y.; Nakano, K.; Kagaawa, M.; Chetani, J.; Mori, Y.; Sunagawa, T.; K
J. Virol. 73, 8053-8063, 1999

A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A and
 A:Reference number: Z22732, MUID:99412319, PMID:10482554

A:Accession: T439a1

A>Status: preliminary, translated from GB/EMBL/TrEMBL

A:Molecule type: DNA

A:Residues: 1 2077 (15%)

A:Cross-references: EMBL:AB021506, NID:94495477, PIDN:AAA78242.1, PID:94960019

A:Experimental source: strain HST, pop. variant B

C:Genetics:

A:Note: U31

C:Superfamily: varicella-zoster virus 240K tegument protein

Query Match 3.0%; Score 124; DB 2; Length 2077,
 Best Local Similarity 20.8%; Pred. No. 3.5;
 Matches 171; Conservative 110; Mismatches 248; Indels 292; Gaps 44;

```

QY 57 SOEYTDVDRSQGSTRKYIYEPFGWKVNNLA-----VEPRNELGSLPLAPEFFPNT 111
DB 584 SEKI,TNTHILRNKFT---KIVOD-----NYNLAAGFMPVTHLNCIFAYL---VNLNPN 633
QY 112 PLLGP-----PPTLOOTEM:-----IKRYGT-HF-----LLSATLGGEESLTIFVDKEL 156
DB 634 EVLKHSGLNINTELLQVEKLYUNTQLFEGISHENINNLNFTLSIRKMEVDFVNSQKP 693
QY 157 SKRARG-----SDNTNSSSVTLETHOLAASYFIDRSTLRRLHHIQIASTAIKV 207
DB 694 SDRASEILAAIESILADPSKNKTIVNIEM:-----KSLEELGKMEISTENKQ 742
QY 208 TETHTGP-----LRSNYNILNLSVSS-----VLWSP-E-----NKIQ 740
DB 743 TARTKQILGDQELTPIYDFLHLHSAYNLPNTTNNLHLHLILKRPDIAALLHDKIQ- 801
QY 241 QGLQVLPTVLYERFV-QAALSYIA'NSRSEFT'KENG'W'HP'PKFPEON'PSMLQAM 299
DB 802 SILDICDDMLNDITVEOTESTVL-----ELV-----DLFPN-----STEKTAL 841
QY 300 FENILRLITETWK--AYNSDPFESDEKLEPMKRLPMNYPLNTSTIMHLWTMSDNQRRYEO 357
DB 842 FESVILTPOLAKK'CANLKTLEPDD-----LAQFITNS-----EQ 877
QY 358 LENSMLQLELKAQIKVHLFLSLSKRCHQKPLISLQPRQSTVYLTFRISQFLYCNENGLLG 417
DB 878 LQNMKKOHF---CKKIPTLMDHILKPLYSQKIITAEKN---WIQR----- 916
QY 418 SFSEETHSCTCPNDVVCTAFLEPCTVETASACLTAPQNKTCWGT'NTGYMLSQCLCKPE 477
DB 917 ---AKTAVITSPPE---LTAFL-----ATAPTKHALQ---TCKPE 947
QY 478 VAPSTDRHYIGFTTNQLEMKYLLQKTDPRPEVHAIFISN--DMRINSWFDPSWKKRMLL 535
DB 948 LDKALOPH-----MEFOMKOTAENDKKHILTPNTLEKPLND-----III 987
QY 536 TLKSNKYK-----SLVHMLGLSLQICLTKNSTLEPV 568
DB 988 ILKNGQSSSLETVHNLNLEPFLKQIQUNDLITPHPHALLPVIKDIETVTSKTSIDLEKI 1047
QY 569 L AYVNPFGGSHSESWEPVNSPFDWERIKLDLPGLCYNNKWTLTGKNW-KTFFETVH 626
DB 1048 LIKTPLNPEQMSKEOKYTPL-----LSFLSKPKKTFECT-- 1082
QY 627 IYLRSRISNCPNGNSIYYEPIEFIDPSRNLGYMKINNIOVFGYSMHFDPPEAIRDLILQ 686
DB 1083 ---EDVKTEIDQMKSTIF--LTKIATSTN-KYTRISH-SVYGQELNLYEERITELKKE 1134
QY 687 LDYPTTGSSJUSALLQ---LLEIKIK-PPVKLSFEPQPPHILPS'LLHPFL----- 742
DB 1135 TNKIKELPSKEYAVAKKKILLSSQDAKTNK-----IYLVNTHHTTKETKNTQPRE 1184
QY 733 ----KLSTSEV----VRIQSAQAFNAKLPNTMDYD-TTKL 764
DB 1185 TASAKALTVEVNNKNEQLQELLNHNFAHLKAKMDQNIHTKL 1225

```

RESULT 6

S36773

GTase-activating protein BMD2 - yeast (Saccharomyces cerevisiae)

N:Alternative names: cyclin CLA2; protein YKL072c; protein YKL424

C:Species: Saccharomyces cerevisiae

C:Date: 03-May-1994 #sequence_revision 03 May-1994 #text_change 29-Oct-1999

C:Accession: S36773; S37917; S37918; S41294; S39117; S42005; S34706

R:Park, H.O.; Chant, J.; Herskowitz, J.

Nature 365, 269-274, 1993

A:Title: BMD2 encodes a GTPase-activating protein for Rad1/Psr1 necessary for proper

A:Reference number: S36773; MUID:93982538, PMID:8371782

A:Accession: S36773

A:Molecule type: DNA

A:Residues: 1-1104 (1A2)

A:Cross-references: EMBL:L19162; NID:9402733, PIDN:AAA34461.1, PID:9402734

R:Pohl, T.M.; Pohl, F.M.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S37897

A:Accession: S37917

A:Molecule type: DNA

A:Residues: 1-681 (30H)

A:Cross-references: EMBL:Z28092; MIPS:YKL092C

A:Experimental source: strain S288C

R:James, C.M.; Gent, M.E.; Indge, K.J.; Oliver, S.G.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S37918

A:Accession: S37918

A:Molecule type: DNA

A:Residues: 237-1104 (1AM)

A:Cross-references: EMBL:Z28092; MIPS:YKL092C

A:Experimental source: strain S288C

R:Chirkova, F.; Nasmyth, K.

EMBO J. 12, 5277-5286, 1993

A:Title: Yeast G(1) cyclins CLN1 and CLN2 and a GAP-like protein have a role in bud t

A:Reference number: S41683; MUID:94085403; PMID:8262070

A:Accession: S41684

A:Molecule type: DNA

A:Residues: 1-436; Y438-1104 (3CV)

A:Cross-references: EMBL:X74130; NID:9395287, PIDN:AAA52228.1, PID:9395289

R:James, C.M.

submitted to the EMBL Data Library, October 1993

A:Reference number: S39116

A:Accession: S39117

A:Molecule type: DNA

A:Residues: 237-1104 (1A2)

A:Cross-references: EMBL:X75561; NID:9414687; PIDN:AAA53241.1; PID:9414689

R:James, C.M.; Gent, M.E.; Oliver, S.G.

Yeast 10, 257-264, 1994

A:Title: Sequence analysis of a 3.5 Kb EcoRI fragment from the left arm of Saccharomy

roten.

A:Reference number: S42004; MUID:94262329; PMID:8203166

A:Accession: S42005

A:Molecule type: DNA

A:Residues: 237-1050 (1A3)

A:Cross-references: EMBL:X75561

A:Experimental source: strain S288C

C:Genetics:

A:Gene: GAP; BMD2; CLA2; ERK25

A:Cross-references: SGD:S0001575; MIPS:YKL092C

A:Map position: 111

C:Superfamily: ras-specific GAP catalytic domain homology

t: 520-741/domain, ras specific GAP catalytic domain homology <GAP>

Query Match 2.9%; Score 118.5; DB 2; Length 1104;

Best Local Similarity 19.3%; Pred. No. 3.5;

Matches 134; Conservative 96; Mismatches 227; Indels 237; Gaps 31;

QY 109 RNRIILGRPTTQQI-----PENIKKVGTHFL-LSATIGGERSLTFV 151

DB 416 KRLKLGKIKITQEIINDTRYNKETRLPIMVDNKNFQIGTICIKISSNLFLPSTNFV 475

QY 152 DKKKLSKRA-----EGSDSTTSSSVTLFTLHOLAASYFIDRSTLRRLH- 198

A: Note: 269-Ser was also found

R: Pohl, J.M.

Submitted to the Protein Sequence Database, May 1996

A: Reference number: S64899

A: Accession: S65272

A: Molecule type: DNA

A: Residues: 1-599 <POW>

A: Cross-references: EMBL: Z73594, NID: g1320500, PDB: e24c950, PDB: g1470501, MIPS: YP243W

A: Experimental source: EMH: strain S28C (AB972)

C: Genes:

A: Gene: SGD: SRP68

A: Cross-references: SGD: S0006164, MIPS: YP243W

A: Map position: 16L

Query Match 2.9% Score 117.5; DB 2; Length 599;

Best Local Similarity 19.9% Pred. No. 1.7;

Matches 72; Conservative 54; Mismatches 120; Indels 115; Gaps 15;

QY 87 NLI AVERRNFGSPPLADEFERNIRILGRPTLOOITENLIKYYG----- 132

Db 168 NNLAL-----VTALEHLKNSLLA-----EEVDNIVNKYQYSLKQYAGNLITTP 213

QY 133 --THFLSATLGE---ESLTFVVK---PKLSKRAEGSDSTNSS----- 170

Db 214 EINNPIVERQSDKDDRLVLLDNGFNMKKITTTSDQKVTTNINWRSFNAKIIDAE 273

QY 171 -SVTLE---TLQLAASYFDROSTLPLR---HHQTASTATKVTETPTGPGGSNYDNL 223

Db 274 VAQFLEGLSIHTQITQYQRLSKLEKALDRHEFFIA-----NIDQ 416

QY 224 LSVSSVLVQSPENKIQGLQVLLVLYQYRQVQAALSVIACNSGPFCKENKQWCHQG 283

Db 317 DDIDEMVENSENN-----QILAYIKYNILLTSISRRDLFTLW 357

QY 284 KPPECNCPMSDQAMEENLLRITETWKAYNSDPDE-----SDEF--KLEMKRLPNNYF 335

Db 358 NQWIKNTSLPSKLIYKQMERIVKNIKYUSDIMELPGVYSDRLLSQDLCKLYQLP 417

QY 336 LNTSTIMILWTMSDFOR-----RYEOLN-----SMKQLFLKAKI--VHKLFSLK 381

Db 418 LNTGTLVLYQSKGYPYMEALALYVAYPPELENKLSIESLDEILLPALLSLNSVRSLOK 477

QY 382 R 382

Db 478 R 478

RESULT 9

T38755

hypothetical protein SPAC305.09c fission yeast (Schizosaccharomyces pombe)

C: Species: Schizosaccharomyces pombe

C: Date: 04-Dec-1999 #sequence_revision 04 Dec-1999 #text_change 03 Dec-1999

C: Accession: T38755

R: Oliver, K., Hartley, D., Wood, V., Barrett, R.G., Rajandream, M.A.

Submitted to the EMBL Data Library, September 1996

A: Reference number: Z21809

A: Accession: T38755

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Residues: 1-2685 <OLI>

A: Cross-references: EMBL: Z99296; PDB: 1CAR16594.1; GSPDB: GNO0066; SDBP: SPAC305.09c

A: Experimental source: strain 972h; cosmid C3H5

C: Genes:

A: Gene: SPOR-SPAC305.09c

A: Map position: 1

Query Match

Best Local Similarity 19.9% Pred. No. 16;

Matches 173; Conservative 124; Mismatches 263; Indels 309; Gaps 52;

QY 75 YKIYREGRMKVN NI AVERFNELG-----SPLPLAPEFRN 110

Db 1455 YDAYDRFSRHYTHCSIALKSPRYLDANAFENVNSYNTIHUTPLVFSHSSWNILFSNM 1514

QY 111 ---IRILGRPTLOOITENLIKYYCTHPTISATLGGPSTTTPV-----PKPKLSK 158

Db 1515 SYPLPSGNLFTPTDSSPN---KKFQPHI---ATPKYALELIPFISIMYKYNFNWMDR 1568

QY 159 RAPGSDSTINSSVILELHQAASYFTRDSILPKPIHH-----QIASIAIKVIFTRI 212

Db 1599 TASATGICAFVNFNFI--LADSEKHEVKNKANLCRKHQFALSMMKVHLARILPKYILNRA 1527

QY 213 -----GPLGCSNYDNLDSV--SSVI-----VQSPNKIQIQLVAV--LILPD 249

Db 1628 ISASIDECAL-----QNSDSTIANVLDDEEKVQCFSPKKNVDCGFAWVDIDYHBAWLLPQ 1682

QY 250 YIQE-----RFVCAALSYIACN--SSGCFPI---CKENKQWCHGCGKPKFPCNC 241

Db 1683 QNEKQSTYPLASPFYV--YHTRKCHPPNFKNFKETIPDTCRGDEFTH-----KCLM 1734

QY 292 PSMD--IQAMFENLIRITETWKAYNSDPDESEKFKPKRLPMNYFELNTSTIMHGLWDSN 350

Db 1735 PSRENPKAVQYELAQ-----KKKALGPFM-----SS 1761

QY 351 FORRY-----EQLE--NSMKQLFLKAKIIVHKLFSLSKRCHQKQPLISLPRTSTYWLTRI 404

Db 1752 RQRRIGFLKSQLSNDSEVRQYERFLKIVILSD--HYRILEYLLKDESSCSQASQ-- 1818

QY 405 QSLFYCNGNGLI---GSPSEPTHS---CICPNQDVVCTAFPLCTVGDASACILCAHD-- 455

Db 1819 -----CSENGQVLSYASLSSESVHAFNNRFVANNVQVKNNF1---RNVMYSYVIEV 1867

QY 456 NKPRGQPNINYSMSQ-----GCKPEVAHSTHYIGPPTDQDLF-----MKYLLOK 503

Db 1868 ERVR---GFAVMSQKALVFLRLDEKRTESANDOFFQNYTETEDRENAHLLKPLLED 1923

QY 504 TDRITEVHAI-----PISNDMLNSWFDSEWPKRM 533

Db 1924 SKRRFWKTSDAIRHSGHSGHNTSISNNFYDIQSIVYFETSPQITQGCCSNF--EKAV 1981

QY 534 LITLSKNYK-----SSLVHMIIGLSL-----QICLKNSLTPVLA--VYVN 574

Db 1987 TISTGSLQIKLSVVLPPIQINVIIFPPPIQVKN--SDEFFSKPSDFVFNASSVLIN 2041

QY 575 PFGGSHSFWMVNFPNFPWEPTKLDLPLQ--GVNLTLLGNSKWKTFPRVHVIYKSR 533

Db 2042 EYGCCHNTVM-----PPW-----VPETTFEDVLPD---AFSRFLH-----RV 2077

QY 634 KSGNPGNESIYIEPLEFIDPSRNLGYMKINNIQVRSYSHIIFDPEAIKOL-----ILQLD 668

Db 2078 -----SFSVIYTKHNDRL--QETIVHSTRAPFEDIDTHADILFD 2115

QY 685 YNYVQGSUSA--LQQLDELKQVKKLSHSGQKKRLKLSGLKHLK 1515E 738

Db 2116 FPPVVFSTSSVYDIFETITTLIYKEFAQF----QPNKPLEFIMLAH--SGLGISE 2170

QY 739 VVEFISALQA-----FNAKLINIMUDYD 762

Db 2171 VVR---ALQARVHRILLDKLQHOL--YDVT 2195

RESULT 10

A44257

interleukin-6 signal transducing molecule gp130 - rat

C: Species: Rattus norvegicus (Norway rat)

C: Date: 30-Apr-1994 #sequence_revision 18 Nov 1994 #text_change 28-Jul-2000

C: Accession: A44257

R: Wang, Y.; Nesbitt, J.E.; Fuentes, N.L.; Fuller, G.M.

Genomics 14, 666-672, 1992

A: Title: Molecular cloning and characterization of the rat liver IL-6 signal transdu

A: Reference number: A44257; M010: 94052397; PMID: 1427893

A: Accession: A44257

A: Status: preliminary; not compared with conceptual translation

A: Molecule type: mRNA

A: Residues: 1-918 <WAN>

A: Experimental source: liver

Query Match 2.8% Score 114.5; DB 2; Length 918;
Best local Similarity 19.2%; Pred. No. 2.3;
Matches 94; Conservative 81; Mismatches 142; Indels 174; Gaps 25;

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-527 <AFO>
A:Cross references: EMBL:AF063866; NID:g4049647; PIDD AAC97719.1; PIDD:g4049759
C:Genetics:
A:Note: MSV240

Query Match 2.8% Score 114.5; DB 2; Length 918;
Best local Similarity 19.2%; Pred. No. 2.3;
Matches 94; Conservative 81; Mismatches 142; Indels 174; Gaps 25;

302 NLPITETWYNSDDESEDFKLPKRLPMNYFLNTSTIMHL...WTMISNTOURVE 456
Db 24 NKLIIIECRNYNLELNIIITYI...DFNPLEIINLDHKNIDGNELKPLE 76
QY 357 QLE-----NSMKQLFLKAKIVHKIFSLKMKCHK...QPLISLP... 493
Db 77 NLEELNSENKLNKLL--PKSIKLL-----NKNINIKDKFLEPLNLEELNISNY 129
QY 394 QPTIYILTRIQSFLYCNENG-----LLGSESETHSTPNDQVVTAFIPVTV 443
Db 130 ISNNSLYCIPKSIKKLNCNCKIKQKFLGGLDNLNLELNISNYDSNLSN LPKSI 187
QY 444 GDSACITTAPONPTPGICNIGYMLSLGKPK-----EVAESTIHYIIGETDLQ 493
Db 188 -----KKLECYCN-----IINCNTLKPENLIELYISITGDSITYY 224
QY 494 DLEMYLLQKTRIFVHAIFINDMKLNSWFFPSWKKPMELKSKNYKSSI VHMIGH 553
Db 225 LQYVYVYVPSVKILKYNNININ-----FPLELENL... 258
QY 554 SIQICLKNSLTBPVLAAYVNPFGSSHSWPMNVNSENFPWERTKILPLQCYNWTIL 613
Db 259 --EIDISNNNNLN--LSLYYIP-----KSVKIL--LKHCHNII 291
QY 614 LNKWKKEFTVHIYFSPFKSNFNGNSIYVEP-----LEFIDPSRNL 658
Db 292 ---DFK-FLELENLIELNTANKNLNLSLYYIPKSVKILFGKGNITDFLEKLK 447
QY 659 GYMKIN---NLOVPGYSMHFDPAIMD ILQIDYPTDGSLSALQLLEIFPWKSLSP 715
Db 348 EILDISNNPNTI--SSLYYIPKSVK--LOYEGCHTD-----FKFLELNLNLELSI 496
QY 716 PGQRRLDLS 725
Db 397 LGNNVNLS 406

RESULT 12
T02275
hypothetical protein T109.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 22-Oct-1999
C:Accession: T02275
R:Vysotskaia, V.S.; Schwartz, J.R.; Kwad, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.;
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel
submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC 11408 sequence.
A:Reference number: Z14649
A:Accession: T02275
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <VYS>
A:Cross references: EMBL:AC004473; NID:g3109236; PIDD:g404972; GSDDB:GN00059; ATSP:11
C:Genetics:
A:Gene: ATSP:T109.12
A:Map position: 114.5; DB 2; Length 454;
A:Introns: 45/114.5; 137.1; 156.3; 213.1; 243.1; 273.1; 303.1; 333.1; 363.1; 393.1; 423.1; 453.1

Query Match
Best local
Matches

2.8% Score 114.5; DB 2; Length 454;
Similarity 21.7%; Pred. No. 2.3;
58; Conservative 40; Mismatches 102; Indels 67; Gaps 11;

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <VYS>
A:Cross references: EMBL:AC004473; NID:g3109236; PIDD:g404972; GSDDB:GN00059; ATSP:11
C:Genetics:
A:Gene: ATSP:T109.12
A:Map position: 114.5; DB 2; Length 454;
A:Introns: 45/114.5; 137.1; 156.3; 213.1; 243.1; 273.1; 303.1; 333.1; 363.1; 393.1; 423.1; 453.1

Query Match
Best local
Matches

2.8% Score 114.5; DB 2; Length 454;
Similarity 21.7%; Pred. No. 2.3;
58; Conservative 40; Mismatches 102; Indels 67; Gaps 11;

QY 427 TCGNDVCTAFLPCTGVDASAGLTGAPDNRTGCTGNTGVM-LSQGLCKPEVAEST--- 482
 DB 173 TLUDGLVETVTPCUGPCTGTCVSSQCCAMNCGCTGPMELKRNLTIAETVPCAVYA 232
 QY 483 ---DHYIGFETDLOLEKMYLLQKTDKRLEVHAIFISNDMKLNSWFDPSWKKRMILTLK 538
 DB 233 RLLSHEVRSITNVPMGMWEPFHNDVNIKAANIMWEN---GLHFSPP--PK---VTVS 284
 QY 539 SNKYKSSIVHMLIGLSIOICITPKNSTPPLVAVYVNPFGGSHSSWMPVNSPNSPDR 598
 DB 285 TSGLVOLKRPKRESKALAVSLNATIDV-----RNIMPIN-----RK 324
 QY 599 TKLMLPLAVYNNWTLITLGNKKWTFEETVHVIYLSPIKSNPNPNESIIYVEPLEFIPSPNL 658
 DB 325 YKLSLLETLRLGLSSKKHYKVLFFVVM-----ACVN-----DSMDARRL 366
 QY 659 GYM-----KINNIOVFGYSMHFDPDA 679
 DB 367 VELVQGIPOCKINLIQ-----FNPIIS 386

RESULT 13
 T41933
 large tegument protein - human herpesvirus 7 (strain J1)
 C:Species: human herpesvirus 7
 A:Variate: strain J1
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T41933
 R:Nicholas, J.
 submitted to the EMBL Data Library, December 1995
 A:Description: Determination and analysis of the complete nucleotide sequence of human
 A:Reference number: 223022
 A:Accession: T41933
 A:Status: preliminary, translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2059, NT
 A:Cross-references: EMBL:U43400; FJUN:AAU54643.1
 C:Genetics:
 A:Note: U31
 C:Superfamily: varicella-zoster virus 240k tegument protein

Query Match 2.8% Score 113; DB 2; Length 2059,
 Best Local Similarity 18.9% Pred: No. 22;
 Matches 150, Conservative 139, Mismatches 248, Indels 216, Gaps 37.

QY 71 FSTRYKTYRPFRCWKVNILAVERRNFLOSLPLAPFERNIRILGRRPTLOQITENI--- 127
 DB 609 FAVLVKLYEDVYSITINGFVEN-----ELIKNI-----EQIYDNTQYL 647
 QY 128 ---KKYGTHTPLSATGCTPSS-ETTPVQPKKISYPAPGSSSTNSSSVTITFTLHQAAS 193
 DB 648 RFLGTRFNMGNITPPTISYVKRWFHDFPLSQKTLIDRAREETINLRFKSVTPFGKQKIATK 707
 QY 184 YFDIPISTPLRHIIQIATAATKVTPTPTGPGGNYD-----NLDSSVSVLVQ 232
 DB 708 NMLREQ--LQGNAMIVVHINIKLITLHQVLSQGLRMILQURLDQLSHNIPSI--- 762
 QY 233 SPINKTQGLGQVLL--IDY--LQRFVQAAISYIACNSIGRFTCKEN----- 276
 DB 763 ---NFVSKLKLHLEKRPPIITLQLEK-VQNILYFYQDLVNEIPAQENSLMTLFIIE 818
 QY 277 ---DWCRCCKPFTNCIS-----MDIQAME--ENLLRITFTKAYNSDPHES-- 320
 DB 819 LPPADSRTH-----ILDTGVYISHRIYKKWLNMRKSLQDAEDLRFINIKRQIGKPFHPG 874
 QY 321 -----DPEKIFMKP-----LPMNYFLNTSTIMILWTMDSNFPQVEQENS 361
 DB 875 KELOKLEKHLIPYKOKVIEYQDVWSEMAKNILITSFSELSPFLASAPTORITQKHKNN 934
 QY 362 MKQ--L--KAKKIVHKIPSTSKS---THKLPPLISLQKQKINISYVWLRPESVYVWNPDL 417
 DB 935 LTKQLLTHMENAKVAMEDECKRFAV--SK---INLEPHILNLIILLKLKLPQFASIGASVLIV 991

QY 418 SFSEEHSHKICENLAVVI---AFLEPTVGLASACLTGAPDNRTGCTGNTGVMLSQGLC 474
 DB 992 ---ENIKLIPPENLIQPSHALL-----SVLIDPFKQKSYSSSEITFFHINPPLE 1041
 QY 475 KPEVAESTHYIGFETDLOLEMK-----YLLQ-----KIDPPIEVHA 512
 DB 1042 TSELLVFKDAYNLKAEFLNALKUSLFATADVUNKADFLIQILDFTYKFRUKTKNGKLLHS 1101
 QY 513 IFISNIMRLNSWFDPSWKKRMILTLKSNKYKSSIVHMLIGLSIOICITPKNSTPPLVAVY 572
 DB 1102 IV-NEDFKLYEETILTELEK-----KATDAKESLTKLKFASEQKIELSRIT---PLAKET 1151
 QY 573 VN---PFGGSHS---ESWEMPVNENSEPDWE-----RIKLIDLPLOCCYNW 610
 DB 1152 LNIETVNPQGVGVVPRPSAFKPAIEVEIKNYPMKI NDLIKFNSHLKTRID-HIQIIN- 1209
 QY 611 TITLGNKKWTFEETVHVIYLSPIKSNPN---GNESIIYVEPLEFI-----DPSPNIGYM 661
 DB 1210 -LSDNKKWDFV-----SKSKISFPPELTISQELIKQPIKIVITEILNKASNDLAY 1260
 QY 662 -----KINNIOV-----FGYSMHFDPDAIRHLLIQIDVPTGSSQSSALLQ 702
 DB 1261 ISEKILKWLIVFKELNTEFFVAINMSEFGEVTFDYKHFPALEYEINSKYIEIKNIQNE 1420
 QY 703 LLEIDRVNKLK 714
 DB 1321 ILENDINIKLS 1332

RESULT 14
 S45734
 probable membrane protein YBL004W Yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein YBL0101
 C:Species: Saccharomyces cerevisiae
 C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 19-Apr-2002
 C:Accession: S45734; S45733; S44556; S37318
 R:Rieger, M.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45734
 A:Accession: S45734
 A:Molecule type: DNA
 A:Residues: 1-2066, RIE
 A:Cross-references: EMBL:Z35765; MIPS:YBL004W
 A:Experimental source: strain S288C
 R:Lohan, A.J.F.; Wolfe, K.H.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45730
 A:Accession: S45733
 A:Molecule type: DNA
 A:Residues: 1214-2493, GLOH
 A:Cross-references: EMBL:Z35765; MIPS:YBL004W
 A:Experimental source: strain S288C
 R:Wolfe, K.H.; Lohan, A.J.F.
 Yeast 10(Suppl.A), S41-S46, 1994
 A:Title: Sequence around the open-reading frame of Saccharomyces cerevisiae chromosome II: si
 A:Reference number: S44556, M01D:94378721; PMID:8091860
 A:Accession: S44556
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1214-2493, <W3L>
 A:Cross-references: EMBL:Z26494; NID:9403311; PID:9403312
 A:Experimental source: strain S288C
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, September 199
 C:Genetics:
 A:Cross-references: Ssp-Sum0100
 A:Map position: 2L
 C:Keywords: transmembrane protein
 F:165-182/Domain: transmembrane #status predicted <IM1>
 F:350-366/Domain: transmembrane #status predicted <IM2>
 F:656-672/Domain: transmembrane #status predicted <IM3>
 F:1021-1037/Domain: transmembrane #status predicted <IM4>
 F:1170-1187/Domain: transmembrane #status predicted <IM5>

membrane *status predicted -TM6>
membrane *status predicted -TM7>
membrane *status predicted -TM8>
membrane *status predicted -TM9>
membrane *status predicted -TM10>
membrane *status predicted -TM11>
membrane *status predicted -TM12>

Query Match Score 112.5% Id 2 Length 2493;

Positive 112; Mismatches 24; Indels 439; Gaps 47;

ESPEWLLSHKGRHGRQAYLVVSRKQVSTPYKIXRF-----80

EVYDAPEV VYNEF ---KTYKDEEMENER 817

VERENFGSPDLADPERNER -----LLGR-----117

VQNVLEKT-----LSKFNENVSALHGLMLVLLGSPNDVQKL 868

QQLLENLKKYGHPLLSALHGLHESLFFVVKRKLKRAE-----161

NKYPNEN-----ELQVLEKELFFELFNGSQSKAEDEKVM 922

-----GSDS 165

APPSGGRSKRAVAVV-----NLSASPRKLNYPFENCH 962

QQLA -----ASYFDKST 191

GRMTSEVNVNLSVLPNFEHENSVAQPIYSTAMAYVLDTEST 1042

AKVTEPRIGHQV-----SNV NLDVSSSVLQV-----S 233

NKROCLKTSVVFVNFELWLSMDIYAVVVKRISHFS 1095

-----LPCVLEDFVQAL-----S 261

FLYWAHNSLYQFYDFEATATAMDTISNOHVKEAVIGPIEPAAD 1155

FLCKENKQWCHGKPEVW-----DSMDIOAMEN-----302

VDLVILIC-----LSVLEKLSLVKLSDSNSTSTPLN 1200

IMKAYNSDFNSDEKPT---FMKPTDMNYFTNST 340

QHVRSFSSLSLKKK-----NLSGLKELALVFNYSWSP 1250

GRTEQ LENSMLLLKAKTVKQFSSFPKHKPLSLFQPTST 498

KIFERNRVSLTEPTEGRKKVFGESKLV--ADLNSYSSSRMHE 1318

INMGLGSSSETHSPFENQVVTAFIPVIGASACTICAPNR 457

ELIDN YKXSE -----LNPVILF-----1349

QGGKKEVAESLHYGFEEVLFEMAYLQKTRTEVHA-----513

-----LFLHFNKKEGLA-----RTNASHALMKFI 1375

QDSWRKEMLLTKPKYKSVVMWVLSGLQTKNSTLEEVLA 572

-----EASKSLSMKKGLFFNIRIGRSLERVSQSFY 1417

GMVNSPDMWRKILKPLQYVWKLQKWKTEFFIVHVLRS 631

EMVKNKYETDFE DMALLYN-----DEEADFTNVNHIQLR 1461

NE SIYYEPL-----EPDGRNLYMKINN-----IQVGYSHFP 675

QGLKNSLSHTLPMTHVYSNPK-----FNIGNELQATGLAQM 1518

QVHTGQSQSAIQLITPRGVNKLSDQAPPLDLSGLLPHPLKL 734

Db 1519 SNNQYKALLPPYISMILKYPNQMKQAVQL-----IVQLSV 1560

QY 735 STSEVVRISQAFNAKAPNTMD 758

Db 1561 ----IVPDGAKSLTSLKPSNLD 1580

RESULT 15

A37181

perforin 1 precursor - human

C:Species: Homo sapiens (man)

C>Date: 31-Jan-1992 *sequence revision 31 Jan 1992 *date change 31 Nov 1999

C:Accession: A43616; A37181; S06432

C:Reichenheld, M.G.; Podack, E.R.

J. Immunol. 143, 4267-4274, 1989

A:Title: Structure of the human perforin gene. A simple gene organization with intere

A:Reference number: A45816; MUID:90070042; PMID:2480361

A:Accession: A45816

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-555 <LIC>

A:Cross-references: GB:M31951; NID:q190339; PID:AAA61367.1; PID:q190340

R:Shinkai, Y.; Yoshida, M.C.; Maeda, K.; Kohata, T.; Maruyama, K.; Yodoi, J.; Yaqita,

Immunogenetics 30, 452-457, 1989

A:Title: Molecular cloning and chromosomal assignment of a human perforin (PFP) gene.

A:Reference number: A37181; MUID:90070333; PMID:2572021

A:Accession: A37181

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-555 <SHI>

A:Cross-references: GB:M28393; NID:q189846; PID:AAA61065.1; PID:q189847

R:Reichenheld, M.G.; Olsen, K.J.; Lu, P.; Lowrey, D.M.; Hamend, A.; Hengartner, H.; P

Nature 335, 448-451, 1988

A:Title: Structure and function of human perforin.

A:Reference number: S06432; MUID:88334728; PMID:3419519

A:Accession: S06432

A:Molecule type: mRNA

A:Residues: 1-331; V' 333-425; S' 427-555 <112>

A:Cross-references: EMBL:X13224; NID:q35385; PID:CAA-1612.1; PID:q35386

C:Genetics:

A:Gene: GDB:PRF1 PFP

A:Cross-references: GDB:118853; OMIM:170280

A:Map position: 10q22-10q22

C:Keywords: transmembrane protein

F1-22/Domain: signal sequence *status predicted <SIG

F1-24-555/Product: perforin *status predicted <MAT>

Query Match 2.7%; Score 112; GB 2; Length 555;

Best Local Similarity 24.0%; Pred. No. 3.8;

Matches 73; Conservative 39; Mismatches 114; Indels 78; Gaps 17;

QY 102 PLAPFPNRIELGPPPP-----LLQQLHNLKKYKPLLSALIGG-----143

Db 188 PLIPDF---KPAIGRIPIHFNASTAPAYLPLISNYSYTHIFPAVELGSPISALIALPTTEL 244

QY 144 -----HESLIIFVOKRKLKRAEGSDSTLNSSSVTFEHLQIAASY---FIDPDS 190

Db 245 AIPQITUNEVEFCTIV---FAQVNIQIGHSISAEAFACEKKEKPKMIASPHQIYPERHS 301

QY 191 TLRLHLLHIIASTAIVETETGFEFSSNPFNLLAVAVLVQVHNFALGLLVLA 1 248

Db 302 EVVGGHHTSINYLPQI---QAGPQYSAWVNSIPSSQGVV---YIEHPHVLHLSQ 353

QY 249 DVIDPFAVAAALSYIAENSEGEFFCKENFWCHGGG---KSPFEN---PSMFI 296

Db 354 DPHPEA PHALSUYLDPA-----PWPTWSPPPGPKSPRPQGVCHSANTTQDC 407

QY 297 QAMEN---LRIT---ETWKAYNSDFESDEFKIPMKKRLPMNPFNLSIMH-----LWTM 347

Db 408 QPPQPG AQLEVFIFQAWGLWGLWETATIDAYVKIFFG-----QLEKISIVWNNNNPIMSV 463

QY 348 DSNE 351

Db 464 RLDF 467
:l

Search completed: May 19, 2003, 10:34:40
Job time : 60 secs

Genome version: 1.1.4 15_4578
Date: (c) 1993 - 2003 Emparean Ltd.

search using SW model

2003, 10:21:44 : Search time 27 Seconds
(without alignments)
1176,699 Million cell updates/sec

us-09-832-129-35

SRAGAEUSMALNEW...GAE-NAKLPNTMDYITIKLS 766

902
10.0 : Gapext 0.5

seps: 41476428 residues

displaying chosen parameters 112892

0

200000000

on March 08

on first 45 summaries

on first 45

number of results per query chance to have a
in or equal to the score of the result being printed,
analysis of the total score distribution

SUMMARIES

Length	ID	Description
587	Q00X85	oncorhynchu
684	Q04779	saccharomye
443	P10643	homo sapien
588	Q9P0V7	paralichthy
1104	P33314	saccharomye
599	P38687	saccharomye
918	P40190	rattus norv
2659	P52162	human herpe
2413	P35194	saccharomye
655	P14222	homo sapien
653	P07228	gallus gall
2259	P35341	simian viru
622	Q9P684	schizosarch
664	P35753	rattus norv
414	Q17290	schizosacch
1041	P36422	tetrahymena
1073	P55204	aus scrofa
574	P06682	oncorhynchu
514	P10820	mus musculus
859	Q06602	homo sapien
207	P30002	human herpe
535	P93703	zea mays (m
441	Q60841	mus musculus
1048	P27895	saccharomye
1442	Q03834	saccharomye
432	Q9P472	xylella fas
1417	Q05214	saccharomye
399	P35739	rattus norv
1042	Q12840	homo sapien
1748	Q00383	epifaqu vi
841	P74397	synecocyst
8726	P20659	drosophila
4568	Q39565	chlamydomon

RESULT 1

ID	COBH_ONCMY	STANDARD	PR1	587 AA
AC	Q90X85			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DI	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Complement component C8 beta chain precursor.			
GN	C8B.			
OS	oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei;			
OC	Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.			
OX	NCBI_TaxID=3022;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
PA	Zarkadis I.K., Kakanizi A., Stryer A.G., Lambiris I.B.;			
RT	*Cloning and characterization of eighth beta component of complement			
RT	in rainbow trout.*;			
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: C8 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX. C8			
CC	BINDS TO THE C5B-7 COMPLEX, FORMING THE C5B-8 COMPLEX. C5-B8 BINDS			
CC	C9 AND ACTS AS A CATALYST IN THE POLYMERIZATION OF C9 (BY			
CC	similarity).			
CC	-1- SUBUNIT: C8 IS COMPOSED OF THREE CHAINS: ALPHA, BETA AND GAMMA.			
CC	THE BETA CHAIN BINDS TO THE C5 ALPHA CHAIN AND TO THE C5B-C7			
CC	COMPLEX, PRESUMABLY TO C5H. IT IS ESSENTIAL TO THE INCORPORATION			
CC	OF C8 INTO THE C5B-C8 COMPLEX (BY SIMILARITY).			
CC	-1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C9, AND TO PERFORIN.			
CC	-1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.			
CC	THIS SWISS PROT entry is Copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by not profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announc/			
CC	or send an email to license@sib.ch)			
EMRL	AF418597; AAL16647.1;			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR002172; LDL_recept_A.			
DR	InterPro: IPR001862; MAC_perforin.			
DR	InterPro: IPR000084; ISPL.			
DP	Pfam: PF00007; ldl_recept_a; 1.			
DP	Pfam: PF01833; MACPF; 1.			
DP	SMART: SM00181; EGF; 1.			
DR	PROSITE: PS00022; EGF_1; 1.			
DR	PROSITE: PS01186; EGF_2; 1.			
DR	PROSITE: PS00968; LDLPA_2; 1.			
DR	PROSITE: PS00279; MAC_PERFORIN; 1.			
DR	PROSITE: PS00092; ISPL; 2.			
KW	Complement pathway; Complement alternate pathway; Glycoprotein;			
KW	Signal; Plasma; Membrane attack complex; Cytolys s; EGF-like domain;			

Q14525 homo sapien
P39346 escherichia
Q08409 kluyveromye
P42504 saccharomye
Q04854 schizosacch
P98157 gallus gall
P98164 homo sapien
Q15423 homo sapien
P02748 homo sapien
Q04778 schizosacch
P24756 schizosacch
P77882 homo sapien

Repeat: Transmembrane.
 KW SIGNAL 1 31
 FT PROPEP 42 44
 FT CHAIN 45 587
 FT DOMAIN 56 112
 FT TRANSMEM 117 159
 FT TRANSMEM 308 324
 FT TRANSMEM 329 348
 FT DOMAIN 495 531
 FT DOMAIN 541 587
 FT DISULFID 118 129
 FT DISULFID 123 142
 FT DISULFID 374 399
 SQ SEQUENCE 587 AA: 65540 MW: 680417659FC879E CRC64;

Query Match 3.7%; Score 153; DB 1; Length 587;
 Best Local Similarity 20.4%; Pred. No. 0.0011;
 Matches 74; Conservative 57; Mismatches 136; Indels 96; Gaps 16;

35 VSDQHATSPD---WLSKRPFF---HPSQVTFDFVPSQGFSTRYKI-----YR 79
 DB 201 IDVREKKPYNQYTLTKGYDKYDQSFYSYSEVHYTWTRSSKTYTVSIGFALPCVA 260
 QY 80 EFG-----RKVNIAVERNEL-----GSPDLAPEFFERNILL 114
 DB 261 EGFENVADSKYSKSEKKIRASRKENSFYQAKAEQLQARYILKREDLMLHPEFFELRAL 320
 QY 115 GRPPTLQQTENLIKKGTHFLLSATLGGESLTIFVUKRKLSKRAEGSDSTNSSVTL 174
 DB 321 PQSYNGEVRQ-LYKDYGTHYTEALGGDYEVTVILKEKLEKTYGSLAYKKEQIVL 379
 QY 175 ETUQLAASYFT-----DPRSTLPL-----IHQIAS-----TAIKVTET 210
 DB 380 KVCANIKGVYTVVGLAGGCGDGLNMGHDIVKGSWMVDYVAVYGGDSFETWLAANKL 439
 QY 211 PTGRL...GCSNYNLDSSVY-----LVQ-----SPENKLLQGLQVLLFDYLERFV 256
 DB 440 PTPLRLMRCEAVHYNLDTRVTPPLVELYVARDPSSANSLK-KNURPALAAYLEE--- 495
 QY 257 QAALSYIACNSGEPTCKENDWCCHGKFPKNCPS-----MDIQAMENILRLITETWA 312
 DB 496 SSSCRAPCRNNGLAVLKGTPTC-----ECVTPSGYSGLGCEITORDPIDIGDSWSC 546
 QY 313 YNS 315
 DB 547 WGS 549

RESULT 2
 YMW5_YEAST
 ID YMW5_YEAST STANDARD: PRT: 684 AA.
 AC Q04779;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAY-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 78.8 kDa protein in ABZ2-CHL12 intergenic region.
 GN YMW075W OR YMW916.14.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SPOUNCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RA Pearson D., Bowman S., Harrell H.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 CC 1- SIMILARITY: TO S. POMBE SPAC16C9.05 AND SPAC2F7.07C.
 CC 1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.ebi.ac.uk/EMBL/nuccore/
 CC or send an email to license@ebi.ac.uk).

CC
 CC
 CC
 DR EMBL: Z48952, CAAB8800.1;
 DR SCD: S0004680; YMK075W.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1
 DR SMART: SM00249; PHD; 2.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS0016; ZF_PHD_2; 1.
 KW Hypothetical protein; Zinc-finger.
 FT 2N_FING 260 309 PHD-TYPE.
 FT DOMAIN 15 30 POLY-SER.
 SQ SPOUNCE 684 AA: 78836 MW: CFC282DC9A782F5C CRC64;

Query Match 3.3%; Score 145.5; DB 1; Length 684;
 Best Local Similarity 18.6%; Pred. No. 0.027;
 Matches 128; Conservative 86; Mismatches 221; Indels 256; Gaps 36;

QY 76 KIVREGRGWKVNLAVERNELGSPDLAP-----EPPNRIKLGKRPDLQOITENLIK 129
 DB 122 KARR-SLWYNNKNIQGSNN SEMTIVSKKPTVFRKFGNPNDAKIFLENMTE 177
 QY 130 KYGTHFLLSATLGGESLTIFVDRKLSKRAEGSDSTNSSVTLTLHQLAASVETDR 189
 DB 178 ESN-----INSTQWQDII-----NPTREPESDQDNKKLSNIEPKILLSINAIY---D 225
 QY 190 SYLRRLHHLQIATAIKVTETHTGPGSGSYNLDSSVTVQSPNKKIQGLQVLLPD 249
 DB 226 SKSKLFGGQNSIKSTS-----NASEKIFEDKNS-----TID 256
 QY 250 YLOERFVQAAISYIACNSGEPFCKENDWCCHGKFPKNC-----PSMDIQAMENILRLIT 307
 DB 257 FINEUPCS-----ACNOSGFLC-----CDTCKSPHFCJLQPPIDPNKPKQDWHON 364
 QY 308 ETWKAYNSDFEESDEFKLMKRLPMNYETINTSTIMILWTDMSNDFQRYEGLNSMKOLEF 367
 DB 305 EC-----KFKIF-----NSMALKKIFESNFK-----UNNNVKIFA 347
 QY 368 KAKIVHKLFSKSKCHKOPILSLPROHSIYWLTRIGSLYCHNGLASFSHSFHSCT 427
 DB 338 KL-----LFNDSHNPKQ--FQUPNYIKETFPVK-----TGRQVSYDE--- 475
 QY 428 GNDQAVVTAPETQVQDASACTICAPNPTGSGFNTGYMLSCFKKFEVAESTHAYIG 487
 DB 376 --NDKIPLT-----DRO--LFNTSY--GQSTIK-----LDSY-N 492
 QY 488 FETDLODLEMVYILAKTDPTPEVIAIFISNIMPLNSWTFPPWPPMLTLKSNFYKSSLV 547
 DB 403 PDTHDSNSGKFLI-----CYKCNQTRASSWHP-NSRLIM----- 434
 QY 548 RMHGLGSLQGLTKKNSLTLEPVIAVYVNPFGSSHSKSFMP VNEHSFDPWERTKLDELPLQ 606
 DB 440 --GLYCVT-----PHLLAVPRASEKN----- 459
 QY 607 CYNWTLTLGNKKW-----TFFETVH-----VLSRSTKSGNGHSIYVEP 648
 DB 460 -----LGSKKCKPLHSHTKVKYKIHHCORSDNSVYKVKKKKRL-----INKNNLYVEP 508
 QY 649 LEFIDPSRLGYMKINNIOVFGYSMHFDPPEAIRDLILQDYPYTGSGDSALLQLEIRD 708
 DB 509 EQ-----KIGYNNNCNIQIPTTSHTDYD-----FNQDFKTIQIDENSIVYDFE---D 553
 QY 709 RVNKLSPQQRKDLDFSCLLHRLKLSISE 738
 DB 554 KLYKSKMWQKKRLFOQESLDKLVNSGQ 583

RESULT 3
 CO7_HUMAN
 ID CO7_HUMAN STANDARD: PRT: 844 AA.
 AC P10643; Q92489;

(1) (Created)
11: Last sequence update
41: Last annotation update
out: 17 precursor.

463:
4: Coordinates: Craniata: Vertebrata: Euteleostomi:
4: Primates: Catarrhini: Hominoidea: Homo.

5: AND PARTIAL SEQUENCE.

pubMed:335508;

akravan D.N., Mueller-Herhard H.J., Fey G.H.:
human complement component C7 and the C5b-7

4354: 540(1988).

FROM N.A.

pubMed:7740625;

to H.A., DISCIPLIN R.J.

human C7 gene and comparison with the C6, C8A C8B

188 514(1995).

AGE SITES.

pubMed:10551889;

ammers M., Hess G., Fendrick A., Miroshnichenko O.;

components of the complement system are

multiple tryptophan residues *;

432786-32794(1999).

21:

pubMed:8871666;

er R., Orren A., Moran B.P., Potter P.C.,

ashina I.V., Shipulin G.A., Lachmann P.J.,

if combined subtotal deficiencies of C6 and C7: their

ation with other C6 and C7 deficiencies *;

948-957(1996).

9.

pubMed:9218625;

A., Sheehan G., Schreiner M., Hobart M.J.;

of C7 deficiency: three different defects *;

019-1026(1997).

220: GLN-482 AND HIS-481

pubMed:9856499;

et M.;

iciency: seven further molecular defects and their

haplotypes *;

4-616(1998).

IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX.

FORMING THE C5b-7 COMPLEX, WHERE IT SERVES

ANCHOR.

MER OR DIMER: AS A C5b-7 COMPLEX IT CAN ALSO

TO ROSETTES.

4 DISULFIDE BRIDGES.

is in C7 are a cause of complement C7 deficiency

redominantly from Neisseria meningitidis.

COMPLEMENT FACTORS C4, C5, C6, AND C7.

NTAINS 2 SUSHI (S) DOMAINS.

NTAINS 2 TSP TYPE 1 DOMAINS.

NTAINS 1 EGF-LIKE DOMAIN.

NTAINS 1 LDL RECEPTOR CLASS A DOMAIN.

try is copyright. It is produced through a collaboration
s Institute of Bioinformatics and the EMBL outstation -
informatics Institute. There are no restrictions on its
of institutions as long as its content is in no way
statement is not removed, made by and for commercial
a license agreement (see <http://www.isb.sib.ch/atomomy>

or send an email to license@sib-sib.ch.

CC EMBL: J03507; AAA51861.1;
CC EMBL: X86328; CAA60121.1;
DR EMBL: X86328; CAA60121.1; JOINED.
DR EMBL: X86330; CAA60121.1; JOINED.
DR EMBL: X86331; CAA60121.1; JOINED.
DR EMBL: X86332; CAA60121.1; JOINED.
DR EMBL: X86333; CAA60121.1; JOINED.
DR EMBL: X86334; CAA60121.1; JOINED.
DR EMBL: X86335; CAA60121.1; JOINED.
DR EMBL: X86336; CAA60121.1; JOINED.
DR EMBL: X86337; CAA60121.1; JOINED.
DR EMBL: X86338; CAA60121.1; JOINED.
DR EMBL: X86339; CAA60121.1; JOINED.
DR EMBL: X86340; CAA60121.1; JOINED.
DR EMBL: X86341; CAA60121.1; JOINED.
DR EMBL: X86342; CAA60121.1; JOINED.
DR EMBL: X86343; CAA60121.1; JOINED.
DR EMBL: X86344; CAA60121.1; JOINED.
DR PIR: A27349; A27340.
DR HSSP: Q07954; ICR8.
DR Genew: HGNC:1346; C7.
DR MIM: 217070;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001884; FacI_MAC.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001862; MAC_perforin.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00084; sushi; 4.
DR Pfam: PF00090; tsp_1; 4.
DR Pfam: PF01823; MACPF; 2.
DR PRINTS: PR00764; COMPLEMENTC9.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00057; FIMAC; 2.
DR SMART: SM00192; ldl_a; 1.
DR SMART: SM00457; MACPF; 1.
DR SMART: SM00209; TSPI; 2.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 1.
DR PROSITE: PS00279; MAC_PERFORIN; 1.
DR PROSITE: PS00092; TSPI; 2.
KW Complement pathway: Complement alternate pathway: glycoprotein;
KW Plasma: Membrane attack complex: Cytolysis; Signal: EGF-like domain;
KW Repeat: Sushi; Transmembrane: Disease mutation.
FT SIGNAL: 1 22
FT CHAIN: 23 843
FT DOMAIN: 24 83
FT DOMAIN: 83 121
FT TRANSMEM: 271 287
FT TRANSMEM: 292 311
FT DOMAIN: 451 487
FT DOMAIN: 497 545
FT DOMAIN: 570 627
FT DOMAIN: 630 689
FT DOMAIN: 695 770
FT DOMAIN: 771 843
FT DISULFID: 85 96
FT DISULFID: 91 109
FT DISULFID: 103 119
FT DISULFID: 337 353
FT DISULFID: 571 613
FT DISULFID: 599 626
FT DISULFID: 631 673
FT DISULFID: 659 688
FT DISULFID: 36 36
FT CAEBQHYD: 292 292
FT CAEBQHYD: 503 503
FT CAEBQHYD: 506 506
TSP TYPE-1 1.
LDL RECEPTOR CLASS A.
POTENTIAL.
EGF LIKE.
TSP TYPE-1 2.
SUSHI 1.
SUSHI 2.
COMPLEMENT CONTROL FACTOR 1 MODULE 1.
COMPLEMENT CONTROL FACTOR 1 MODULE 2.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
N-LINKED (MAN).
N-LINKED (GLCNAC...) (POTENTIAL).
C-LINKED (MAN) (PARTIAL).
C-LINKED (MAN) (PARTIAL).

```

FT CARBOHYD 509 C-LINKED (MAN) (PARTIAL).
FT CARBOHYD 754 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARIANT 220 R -> Q (IN C7D).
FT /FTID-VAR_012643.
FT VARIANT 379 G -> R (IN C7D).
FT /FTID-VAR_012644.
FT P -> S (IN C7D).
FT VARIANT 521 /FTID-VAR_012645.
FT E -> Q (IN C7D).
FT VARIANT 682 /FTID-VAR_012646.
FT R -> H (IN C7D).
FT VARIANT 687 /FTID-VAR_012647.
FT R -> V (IN REF. 2).
FT CONFLICT 152 R -> T (IN REF. 2).
FT CONFLICT 389 P -> T (IN REF. 2).
FT CONFLICT 587 P -> T (IN REF. 2).
FT CONFLICT 821 CA -> A (IN REF. 2).
FT CONFLICT 843 MW: DA7817C92E231FA5 CRC64.
SQ SEQUENCE 843 AA: 93514 MW: 3.1% Score 126; DB 1; Length 843;
Query Match
Best Local Similarity 20.4% Pred. No. 0.18.
Matches 107; Conservative 57; Mismatches 153; Indels 208; Gaps 29;
QY 32 VAASVQHAISPPDWLISKGGPHK-----SDEYDFVIRSKGQSPSHYKIRPFGKWKV 86
Db 208 VKHTSTEHTSS-----SRFRSFFSSSSSSPSYTSHTNEIRKSKSYLLV --V 253
QY 87 NNIAVPRPFLQSP---LPIADPPFNPIIGPPTIQQIT--FN: IKKYGPHPLISATI 141
Db 254 EN-TVEAQP-INNNPFIQLAPFNKE---LSHPLSLDYSAIRRLIDQYCHYLQSGSL 309
QY 142 GTEESITIVQPKLSKRAESSESTNSSVLTLEHLGLASYSIDEDSTLRRLHILQIA 201
Db 310 GQYKVIYFYDPSFKIQ-----NIFNSVEKKCKSSQWHFVKPSS---H----- 351
QY 202 STAIKVTETRTGPGGSGNYDNLSSVSVLQSPENKIGLQGLQVLLPYIQERFVQ--- 257
Db 352 -----GCKELNALKKAAS-----GTNNV-LRG-----EPPIKRGGA 382
QY 258 ---AALSYIAC-NSRDEPFICKKCNKCHGCKPKPKPCNCPMDIQAMPNLIPIETWKAY 313
Db 383 GFTSGISYIFLDNPAGN--KRRYSAAWASVNIIP-----QVTKKQIPHYE----- 426
QY 314 NSDFEESDEKFLMKPLPNYELNTSTIMLTMDSNFQRYEQLNSMKOLFIAKQIV 373
Db 427 -----LVKIVPC-----ASVKKLYIK--WAL 445
QY 374 HKLPSLSKSKCHKOPLISLPORSTSYWITRIQSPLYCNGHIGLGSFSETHSCPCNDQV 433
Db 446 EYVLDEFDPCHCRP-----CQNGSL--ATVEGTH--CLAH--- 476
QY 434 VCTAFILPCTVGDA-----SACILTCAPDNRTRCCGTCGTCGMYLSQG 472
Db 477 -CK---PYTFGACPGGVLVGACGVGSGWSWSVCGGKKTRSRCCNPPPSGGG 532
QY 473 -LCKPEVAESTDIHYIGFTDLQLEMKYLLQKTRDPIEVHIAFTIS 516
Db 533 RSCVGHETTES-----TOCHDELEHL-----KLEPHCPPLS 564
RESULT 4
ID C08B_PAROL STANDARD; PRT: 588 AA.
AC Q9PMW7;
DT 15-JUN-2002 (rel. 41; Created)
DT 15-JUN-2002 (rel. 41; Last sequence update)
DE 15-JUN-2002 (rel. 41; Last annotation update)
DE Complement component C8 beta chain precursor
GN C8B.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Paralicthidae;
OC Pleuronectoidae; Paralicthiidae; Paralicthys.

```

```

OX NCBI_TaxID-8255;
RN [1] SEQUENCE FROM N.A.
RX MEDLINE-20009525; PubMed-10541805;
RA Katagiri T., Hirono I., Aoki T.;
RT "Molecular analysis of complement component C8beta and C9 cDNAs of
PT Capraes flounder, Paralichthys olivaceus.";
RL Immunogenetics 50:43-48(1999).
CC 1- FUNCTION: C8 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX. C8
CC BINDS TO THE C5B-7 COMPLEX, FORMING THE C5B-8 COMPLEX. C5-B8 BINDS
CC C9 AND ACTS AS A CATALYST IN THE POLYMERIZATION OF C9 (BY
CC similarity).
CC 1- SUBUNIT: C8 IS COMPOSED OF THREE CHAINS: ALPHA, BETA AND GAMMA.
CC THE BETA CHAIN BINDS TO THE C8 ALPHA CHAIN AND TO THE C5B-7
CC COMPLEX, PRESUMABLY TO C5B. IT IS ESSENTIAL TO THE INCORPORATION
CC OF C8 INTO THE C5B-C8 COMPLEX (BY similarity).
CC 1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C9, AND TO PERFORIN.
CC 1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC 1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (s.a. http://www.isb.scripps.edu/
CC or send an email to license@isb-sib.ch).
EMBL: AB020962; BAA06877.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR031862; MAC_perforin.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00057; ldl_recept_a; 1.
DR Pfam: PF01823; MACPF; 1.
DR Pfam: PF0764; COMPLEMENTC9.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00192; LDLA; 1.
DR SMART: SM00457; MACPF; 1.
DR SMART: SM00209; TSP1; 2.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01209; LDLA_1; 1.
DR PROSITE: PS00068; LDLA_2; 1.
DR PROSITE: PS00279; MAC_PERFORIN; 1.
DR PROSITE: PS00092; TSP1; 2.
KW Complement pathway; Complement alternate pathway; Glycoprotein;
KW Signal; Plasma; Membrane attack complex; Cytolysis; EGF like domain;
KW Repeat; Transmembrane.
FT SIGNAL 1 30 POTENTIAL..
FT PROPEP 31 46 HY SIMILARITY.
FT CHAIN 47 588 COMPLEMENT COMPONENT C8 BETA CHAIN.
FT DOMAIN 58 114 TSP TYPE-1.
FT DOMAIN 115 152 LDL-RECEPTOR CLASS A.
FT TRANSMEM 408 324 POTENTIAL..
FT TRANSMEM 329 348 POTENTIAL..
FT DOMAIN 495 531 EGF-LIKE.
FT DOMAIN 542 588 TSP TYPE-1 2.
FT DISULFID 118 129 BY SIMILARITY.
FT DISULFID 123 142 BY SIMILARITY.
FT DISULFID 146 151 BY SIMILARITY.
FT DISULFID 474 399 HY SIMILARITY.
SQ SEQUENCE 588 AA: 65872 MW: 0C525HR95E7AC2AF CRC64;
Query Match
Best Local Similarity 3.0% Score 122; DB 1; Length 588;
Matches 71; Conservative 53; Mismatches 114; Indels 88; Gaps 15;
QY 46 WLLSKKQPP---HSSQYTHVWVSGQSPSHYKIRPFGKWKV-----YEPFGWVNNIAVER 93
Db 210 YTLTFRTYFLVNVQSFESYSQYVYSMEPEPMTIVLHAIPLIAEPF INYNSAKVTE 279

```


* Substitutions as indicated in no way
statement is not removed. Sage by and for commercial
license agreement (See <http://www.isb-sib.ch/annouche/>
to license@isb-sib.ch).

```

196 100 ANNOTATED_ORF.
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

QY 402 TRQSLYCNENCLL-----CSSEETHSC-----DYNQVAVVIA 447
Db 382 NGTELYNLNNRYVASLAARNVVKSPATVLTIPGSHFKASH VDLKAFKDLJAVE 441
QY 438 FLDP-----CTVGDASACLTICAPDNKTRGTCNTRYM--LSQGLA-- 474
Db 442 WTPPSKPVNKYULEWCVLSNSPCT---DWQDQDCTVNRHFKGSLLESKYLTITVTV 498
QY 475 ---KPEVAESTDIHYIGFETDLDQLEMK-----YLLQKTDRFEVIAIFISNMRLNS 523
Db 449 FQDGPSPSPSMKAYIKCAAPSKGPTVPTFKVGNVAVIEWHGFVIVGNCPTEN 552
QY 524 WPDPSWR---KRMILTILKSNKYKSSLVHMLIGLSIQICLTKNSTLEFV LAVVYNPQG 578
Db 553 -YSISVTSVSGKEMVVRVDSSTHTXTLSL-----SSDTLYVWUAAAYTEERK 600
QY 579 SHSESFWPVNENSFPDWRKTKIDLPQCYNWLT---LGNKKKIFEFVHIVSRKSN 636
Db 601 DGPEFTFTTL---KPAQGEIAIVVPV-GLAFITLILG---VEGFNKRHFKHFWPN 653
QY 637 GPNESIIYEPLEFIDPSRNLGYMKINNOVEGYSMIIFDEAIRELLIQLIYPTQSG 696
Db 654 VP-----DPSKS-----HIAOWSPHPP KHNNSKLOMYSDANF 687
QY 697 DSALLQILFIRHVNKLSPPGQRRIIDF-----SCLRHRLKLSIE 738
Db 688 TD--VSWEIEANNNKPKPCDLKSLTLFKKFKISTEGRSSQ---SSWSSAPP-ESSE 744

RESULT 8
TEGU_HSV7J STANDARD: PRT: 2059 AA.
AC F52362;
DI 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
BT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Large tegument protein.
GN 031.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN 11.
RP SEQUENCE FROM N.A.
RA Nicholas J.
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV 1 UL36,
CC EHV-1 24, EBV BFLP1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouche/
CC or send an email to license@isb-sib.ch)
CC
CC PMR1: 043400; AAC54693.1;
DR
SQ SEQUENCE 2059 AA: 239476 MW: AAE6CA60DEDC5D10 CRC64;

Query Match 2.8%, Score 113, EB 1, Length 2059;
Best Local Similarity 18.9%, Pred No. 5, 5;
Matches 150; Conservative 128, Mismatches 298, Indels 216, Gaps 37;

QY 71 FSTRVK YPEFGKWKVNNIAVPRFPSPLOPLAEFFRNILGRPTIIOGTIENI--- 127
Db 609 FAVLKLVDVTSITINGEVEN-----ELIKNI-----EIVDNTQYL 647
QY 128 -IKK*GTHFLLSATLGEES ITIFVEKPKLSPAPASSTINGSCVTLTHLQAS 183
Db 648 KQGTNNKNTPTTISVFKWFLDEFSQTLIIIPAFIENDEFSVDFGKQIAIK 707

```


RT RT Localization and molecular modelling of the membrane-inserted domain
RL RL of the ninth component of human complement and perforin.
[6]
RP VARIANTS HP.H2 G-183; W-225; S-252; Y-279; L 445 AND E 429.
RX MEDLINE-20030957; PubMed-10581959;
RA Stepp S.E., Dufourcq-Iagourose R., Le Beist F., Bouwan S., Certain S.,
RA Mathew P.A., Hentger J.-L., Hennett M., Fischer A., de Saint Basile G.,
RA Kumar V.;
RT *Perforin gene defects in familial hemophagocytosis
RT Lymphohistocytosis.*
RL Science 286:1957-1959(1999).
[7]
RN VARIANTS HP.H2 MET-50; ASP 224 AND LYS 285 DEL.
RX MEDLINE-21030530; PubMed-11179007;
PA Goerandaecker E., Samuelsson A., Janka G., Schneider M., Giersey A.,
PA Soederhaell G., Samuelsson A., Janka G., Schneider M., Giersey A.,
PA Valman N., Xerost T., Eickler R., Jahnkeinen K., Strom Mathiesen L.,
RA Haraldsson A., Poole J., de Saint Basile G., Nordenskjold M.,
RA Hentger J.-L.;
RT *Spectrum of perforin gene mutations in familial hemophagocytic
RT Lymphohistocytosis.*
RL Am. J. Hum. Genet. 68:560-597(2001).
CC -1- FUNCTION: IN THE PRESENCE OF CALCIUM, PERFORIN POLYMERIZES INTO
CC TRANSMEMBRANE TUBULES AND IS CAPABLE OF LYSING NON SPECIFICALLY
CC A VARIETY OF TARGET CELLS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC
CC T-LYMPHOCYTES.
CC -1- INDUCTION: Repressed by contact with target cells.
CC -1- DISEASE: DEFECTS IN PRF1 ARE A CAUSE OF FAMILIAL HEMOPHAGOCYTIC
CC LYMPHOHISTOCYTOSIS TYPE 2 (HHL2 OR FHL). A RARE AND LETHAL
CC AUTOSOMAL RECESSIVE DISORDER OF EARLY CHILDHOOD CHARACTERIZED BY
CC EXCESSIVE IMMUNE ACTIVATION. PATIENTS HAVE A NONMALIGNANT
CC ACCUMULATION AND MULTIVISCERAL INFILTRATION OF ACTIVATED T
CC LYMPHOCYTES AND HISTIOCYTES (MACROPHAGES).
CC -1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C8, AND C9.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
EMBL; M1951; AAA60167.1; ;
DR EMBL; X13221; CAA31612.1; ;
DR EMBL; M28393; AAA60065.1; ;
DR EMBL; L40557; AAA61618.1; ;
DR PIR; S06432; S06442;
DR PIR; A37181; A37181;
DR PIR; A45816; A45816;
DR Genew; HGNC:9360; PRF1;
DR MIM; 170280; ;
DR MIM; 603553; ;
DR InterPro; I3R000008; C2;
DR InterPro; I3R0000561; EGF-like.
DR InterPro; I3R001862; MAC_perforin.
DR Pfam; PF00138; C2; 1;
DR Pfam; PF01823; MACPF; 1;
DR SMART; SM00239; C2; 1;
DR SMART; SM00457; MACPF; 1;
DR PROSITE; PS00279; MAC_PERFORIN; 1;
DR PROSITE; PS00222; KCF_1; FALSE_NFG.
DR PROSITE; PS01186; EGF_2; FALSE_NFG.
DR PROSITE; PS04499; C2_DOMAIN_1; 1;
DR PROSITE; PS00044; C2_DOMAIN_2; 1;
KW Glycoprotein; Transmembrane; Cytolysis; Calcium-binding; Signal;
KW EGF-like domain; T-cell; Disease mutation; Polymorphism.
FT SIGNAL 1 21 555 PERFORIN 1.
FT CHAIN 22 555 PERFORIN 1.

```

FT TRANSMEM 188 204 POTENTIAL..
FT TRANSMEM 212 231 POTENTIAL..
FT DOMAIN 373 408 EGF-LIKE.
FT DOMAIN 416 498 C2 DOMAIN.
FT DSULFID 257 279 BY SIMILARITY.
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 549 549 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 50 50 V -> M (IN HPLH2).
FT VARIANT 123 123 /FTID=VAR_010772.
FT VARIANT 123 123 R -> H.
FT VARIANT 183 183 /FTID=VAR_010773.
FT VARIANT 183 183 V -> G (IN HPLH2).
FT VARIANT 224 224 /FTID=VAR_010744.
FT VARIANT 224 224 I -> D (IN HPLH2).
FT VARIANT 225 225 /FTID=VAR_010774.
FT VARIANT 225 225 R -> W (IN HPLH2).
FT VARIANT 252 252 /FTID=VAR_010745.
FT VARIANT 252 252 N -> S (IN HPLH2).
FT VARIANT 279 279 /FTID=VAR_010746.
FT VARIANT 279 279 C -> Y (IN HPLH2).
FT VARIANT 285 285 /FTID=VAR_010747.
FT VARIANT 285 285 MISSING (IN HPLH2).
FT VARIANT 345 345 /FTID=VAR_010775.
FT VARIANT 345 345 P -> L (IN HPLH2).
FT VARIANT 429 429 /FTID=VAR_010748.
FT VARIANT 429 429 G -> E (IN HPLH2).
FT CONFLICT 332 332 /FTID=VAR_010749.
FT CONFLICT 426 426 L -> V (IN REF. 2).
FT CONFLICT 426 426 G -> S (IN REF. 2).
SQ SEQUENCE 555 AA; 61377 MW; DDFDEODICAR7586R CRC64,

Query Match 2.7%; Score 112; DB 1; Length 555,
Best Local Similarity 24.0%; Pred. No. 1.1;
Matches 73; Conservative 39; Mismatches 114; Indels 78; Gaps 17;

QY 102 PLAFERNIRIGRPP-----TLOQITENIKKYCTHEILLSATLGG----- 143
Db 188 PHPPK---KRALGDLPHFNASTOPAYLRLLISNYGTHFIRAVELGGKISALTRICEL 244
QY 144 -----EESLTFVDKRLKRAEGSDSTTSSSVTLLEHQAASY---FIDRDS 190
Db 245 ALRGITNEVEELUVP---KQVNIQIHCSISAKACAKKKKKHKKMTASHPHQYRKHS 301
QY 191 TLRRLHIOIASTAKVETRTGPGICSNYDINDSVSVVQSPNKTQLOGLQVLL--P 248
Db 362 EVWG--HHTSNILLRPI---QASPEYSANVNSLPGSPGLVD-----YTLEFLHVLDSQ 353
QY 249 DYLGQERFVQAALSYIACNSRGEFTCKENKWCWCHQGP---KPEP--CNC-----PSMDI 296
Db 354 DPREALRPALSYLTDRH-----RWRDCSRPCPPGQKSPKDPCCVCCHCSAVTTQUC 407
QY 297 QAMEENL LRIT--ETWKAYNSDFEESDEF-KLPMKRLPMNYFLNTSTIMI---LWTM 347
Db 408 QPQRCGLAQLEVFTEQAWGLWETWPTATDAVYKTFPG---QQLPFTSTVYVNNNNPQWSV 453
QY 348 DSNF 351
Db 464 PLDF 467

RESULT 11
ITBL_CHICK
ID ITBL_CHICK STANDARD. PPT. 803 AA.
AC P07228;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Integrin beta-1 precursor (CSAT antigen) (JC22 antigen) (RGD-
receptor).
GN ITGB1.
OS Gallus gallus (Chicken).
OC Fukuayota, Metazoa; Chordata, Craniata, Vertebrata, Euteleostomi;
OC Archosauria, Aves, Neognathae, Galliformes; Phasianidae; Phasianinae;

```

```

OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Embryonic fibroblast;
RA Tamkun J.W., Besimone D.W., Fonda D., Patel P.S., Buck C.,
RA Horwitz A.F., Hynes R.O.;
RT *Structure of integrin, a glycoprotein involved in the transmembrane
RL Linkage between fibronectin and actin";
CC FUNCTION INTEGRINS ALPHA 1/BETA-1, ALPHA 2/BETA-1, ALPHA 3/BETA-1,
CC 1 AND ALPHA-11/BETA-1 ARE RECEPTORS FOR COLLAGEN. INTEGRINS ALPHA-
CC 1/BETA-1 AND ALPHA-2/BETA-1 RECOGNIZE THE PROLINE-HYDROXYLATED
CC SEQUENCE G-P-P-G-E-P IN COLLAGEN. INTEGRINS ALPHA 2/BETA-1, ALPHA-
CC 3/BETA-1, ALPHA-4/BETA-1, ALPHA 5/BETA-1, ALPHA 6/BETA-1, ALPHA
CC 10/BETA-1, ALPHA-11/BETA-1 AND ALPHA-V/BETA-1 ARE RECEPTORS FOR
CC FIBRONECTIN. ALPHA-4/BETA-1 RECOGNIZES ONE OR MORE DOMAINS WITHIN
CC THE ALTERNATIVELY SPLICED CS-1 AND CS-5 REGIONS OF FIBRONECTIN.
CC INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRINOGEN. INTEGRIN
CC ALPHA-1/BETA-1, ALPHA-2/BETA-1, ALPHA-6/BETA-1 AND ALPHA-7/BETA-1
CC ARE RECEPTORS FOR LAMININ. INTEGRIN ALPHA-4/BETA-1 IS A RECEPTOR
CC FOR VCAM1 AND RECOGNIZES THE SEQUENCE Q-1-D-S IN VCAM1. INTEGRIN
CC ALPHA-9/BETA-1 IS A RECEPTOR FOR VCAM1. CYTOACTIN AND
CC OSTEOPONTIN. IT RECOGNIZES THE SEQUENCE A-E-I-D-G-I-E-I IN
CC CYTOACTIN. INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR EPILIGRIN
CC AND THROMBOSPONDIN. INTEGRIN ALPHA-V/BETA-1 IS A RECEPTOR FOR
CC VITRONECTIN. BETA-1 INTERFERS WITH THE SEQUENCE R-G-D IN A
CC WIDE ARRAY OF LIGANDS.
CC -1- SUBUNIT: HEHEKQIMK OF AN ALPHA AND A BETA SUBUNIT. BETA-1
CC ASSOCIATES WITH EITHER ALPHA-1, ALPHA-2, ALPHA-3, ALPHA-4, ALPHA-
CC 5, ALPHA-6, ALPHA-7, ALPHA-8, ALPHA-9, ALPHA-10, ALPHA-11 OR
CC ALPHA-V.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www1.sib.scrib.ch/announcements
CC or send an email to license@sib.scrib.ch).
CC -----
CC EMBL, M14049, AAA48926.1, .
CC PIR: A23947; IJCH3.
CC HSP: P05106; IJCH3.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002369; Integrin_H.
CC InterPro: IPR001169; Integrin_beta_C.
CC InterPro: IPR003659; Plexin-like
CC InterPro: IPR002035; VWFA.
CC Pfam: PF00362; Integrin_B; 1.
CC PRINTS: PR01186; INTEGRIN.
CC ProDom: PD001811; Integrin_B; 1.
CC SMART: SM00187; INB; 1.
CC SMART: SM00423; PSI; 1.
CC SMART: SM00327; VWFA; 1.
CC PROSITE: PS00243; INTEGRIN_BETA; 3.
CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
CC Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Phosphorylation; Signal.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 803 INTEGRIN BETA-1.
FT DOMAIN 25 733 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 734 756 POTENTIAL.
FT DOMAIN 757 803 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 144 382 VWFA-LIKE.
FT DOMAIN 471 640 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 471 520 1.
FT REPEAT 521 564 11.
FT REPEAT 565 603 111.

```


QY 202 STAI-----KVTEET----- 212
 Db 1076 QSSVKKKKKQGGGHTTPTIMKSPFKKIDSKPTILSVENINVLAVNENLIHHPNPDP 1135
 QY 213 QDQAGSNVNMID-----SVSSVLQSPENKIQACIQVILP----- 248
 Db 1146 GYLQVITITNCVSDIAEPLKLSWSSL-----NRTLEHETPDPIEVNVCALVIGVCE 1190
 QY 249 DY-----LQRRVVOALSYI----- 263
 Db 1191 CQYCMGCDTFTWFFLPKQJHFGDQPKNPDPVYVQSKTPEKRV-ASMAYVKCATSSL 1249
 QY 264 --A--NSEET--KENA--W--H--G--K--P--E--N--S--M--D--J--A--M--E--N--L--R--T--E--T--W--K--A--Y--N--S--D--F--E--S--D 321
 Db 1250 KAALRGAGVFWAYGVDNDANNDIADIAN-----TPVQISKPOLQALPTIPTSANIT-HRLD 1305
 QY 322 EFKLPKRLPMNFTINTIMHWTWDSNFORRYKOLNSKQPLKAKKIVHKLPFLSK 381
 Db 1406 DGATTKFTPASSYAFSS-----YTHISNDQ-QYIEVDPVVDNIIYQQLMITGLGIE 1359
 QY 382 KCKQPLISLPKORTSYWITRIQSPLYCNECLGSESETHSCIGNDQVCTAFILPC 441
 Db 1360 TYNPPV-----PTSEELT-----LHETSS-----C 1383
 QY 442 TVIDASALITAPDNTPTGNTNTNYMI SQGLAKPEVAESTDHYIGFETRLQPLPMKYL 501
 Db 1384 CTRPVPVQITCE-----SPSLLPQITVHYENPVDYDNPADYEL----- 1423
 QY 502 QKTDPRFEVHAIFISNIMPLNSWFOISWKKRMLITIKSNKYKSSIVHMLIGLSIQCLTK 561
 Db 1424 -----AILDYLSQARIGSTEHYSLQDR--IDLLAHLTAKOMINSIIGLDESVSLLN 1473
 QY 562 NSTHEPVLAVVNPFGGSHSEPMFVNPENSPDWERTKIDILPQCYNWKTLTLG----- 615
 Db 1474 DAV---VTSQYTN-----NW---ISPCSV-----TKDILVFKMMANNILHLSHPQY 1514
 QY 616 ---NKWKTEFEVHILYLSRYSKSPNGNESIYEP-----LEPIDSR-----NL 658
 Db 1515 YLRITWNSIFDYTWYR-P-RPGNAITNIAATLSHPKLLPRAMNIDWTPVHAIVLASL 1573
 QY 659 GYMKINNVQFYSMHDPDEAFPIILQIDYDPTQGSQNSALLOLELDRVKNKLSIPQO 718
 Db 1574 DYIKL-SIDALQWCI---KQVLADLUNGIDYELLILSENS-----IELSDPAMNLI---A 1621
 QY 719 RRLDLSCLLRHKL 734
 Db 1622 RKLTTLAALQCNQGLVL 1637

RESULT 13

ME25_SCHPO
 ID ME25_SCHPO STANDARD; PRT: 622 AA.
 AC Q9P6S4; O96WS0; P78952;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Molecular expression, upregulated protein 25.
 GN ME25 OR SHC27.04.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales, Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID:4896;
 RN 111
 RP SOURCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE-21848401; PubMed-11859360;
 RA Wood V., William R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
 RA Soudros J., Peat N., Hayes J., Baker S., Hasham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor K., Cronin A., Davis P., Feldwell I., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,
 RA Mendley P., Meade S., Mungall K., Murphy L., Nibbeli D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whithead S.,
 RA Woodward J., Wal-kerti G., Aert P., Eklund J., Graymoffe B.,
 RA Welljens I., Vanstreele E., Rieger M., Schaefer M., Mueller Aier S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,
 RA Horzym K., Langer I., Beck A., Ichirak H., Reinhardt K., Pohl T.M.,
 RA Iger T., Zimmermann W., Wedler H., Wambolt R., Purnelle H.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galliani F., Aves S.J., Xiao S., Hunt C., Meade K., Bust S.M.,
 RA Lucas M., Pochet M., Gaillardin C., Tallada V.A., Garcon A., Thode G.,
 RA Baga P.R., Cruzado L., Jimenez J., Sanchez M., Del Rey P., Benito J.,
 RA Dominguez A., Revucita J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.K., Paulsen O., Forsburg S.L.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 PT "The genome sequence of Schizosaccharomyces pombe.";
 PL Nature 415:871-880(2002).
 RN 121
 RP SOURCE OF 339-606 FROM N.A.
 RC STRAIN-CD18-1;
 RX MEDLINE-21270454; PubMed-11376151;
 RA Watanabe T., Miyashita K., Saito T.T., Yoneki T., Kakihara Y.,
 RA Nabeshima K., Kishi Y.A., Shimoda C., Nojima H.,
 RT "Comprehensive isolation of nucleus specific genes identifies novel
 RT proteins and unusual non-coding transcripts in Schizosaccharomycetes
 RT pombe.";
 PL Nucleic Acids Res. 29:2327-2337(2001).
 RN 131
 RP SOURCE OF 417-622 FROM N.A.
 RC STRAIN-972;
 RA Kohno S., Niwa O., Yano M., Saitoh S., Katayama I., Nagao K.,
 RA Yanagida M.,
 RT "S. pombe chromosome II cosmid 1228 sequence.";
 PL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/bdai/doc/>)
 CC or send an email to license@ebi-sib.ch.
 DR EMBL; AL353866; CABR9003.1;
 DR EMBL; AB054309; BAB60876.1;
 DR EMBL; D83993; BAAL2202.1;
 KW Melosis.
 FT CONFLICT 577 577 V -> A (IN REF. 3).
 FT CONFLICT 608 622 NKVAFIDNEALLYOV -> FUSRI (IN REF. 3).
 SQ SEQUENCE 622 AA: 71463 MW: 10486488AC98676 CRC64;
 Query: Match 2.78, Score 11.0, DB 1, Length 622;
 Best Local Similarity 18.58, Pred. No. 1.8;
 Matches 107; Conservative 64; Mismatches 155; Indels 252; Gaps 26;
 QY 29 VLAAVAVSDQHAISPFQWILSDKGPFRHSQRYTFVDRKROCFSTRYKTYRFGKWKVN 88
 Db 173 VLALRAISSVDEENKRELSPSDPVTESVYVDNGELSNQ-----KVET 217
 QY 89 LAVERNFILGSPILAPPE--FRNRIILGRPTIQQITEN-----LKKYGTPL 146
 Db 218 VEISEKQNDPDAGYSQVILLDFKSKCKOKKVITLPSIFENEGTFPHDITFMTDLGTPY 277
 QY 137 LSATLAGEESLTIFFVQPKKISKAFSGSTNTSSSVLTLEHOLAASY-FIDRSTLREL 195
 Db 278 ---SIGNERFKQILDVGLLANDE-----NLVIGTLAEYCFMFDSDLIQ 323
 QY 146 HHQIATAIKVITPTTGTGLOCSNYNDLSVSSSVLVQSPENKIQGLQVILPEVIGERF 255
 Db 324 R-----NVEHTEKELSLV-----QFKG----- 340

RA Yamamoto A., West P.R., McIntosh J.P., Hiraoka Y.,
 RL Submitted (AUG-1997) to the EMBL/GenBank/INRJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE-21848401; PubMed-11859460;
 RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mundell K., Murphy I., Nibbelin D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
 RA Rutherford K., Putter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkov G., Ait P., Robben J., Grymonprez B.,
 RA Wolfjens J., Vansteels E., Kieger M., Schaefer M., MacIver Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambuit R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Kocher M., Galland J., Tallada V.A., Garza A., Thode C.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta I.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Polashkin J.,
 RA Shpakovski G.V., Ussery D., Harrell B.G., Nurse P.,
 RT "The genome sequence of *Schizosaccharomyces pombe*,"
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
 CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
 CC ORGANELLES ALONG MICROTUBULES. REQUIRED FOR NUCLEAR MOVEMENT
 CC DURING MEIOTIC PROPHASE.
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
 CC INTERMEDIATE AND LIGHT CHAINS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROBABLY BINDS INDIRECTLY TO
 CC THE INNER PLASMA MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AB006784; BAA22056.1;
 CC EMBL: AL355652; CAB90788.1;
 CC EMBL: AL352839; CAB60251.1;
 KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
 FT DOMAIN 1217 1252 COILED COIL (POTENTIAL).
 FT DOMAIN 1984 2012 MICROTUBULE-BINDING (POTENTIAL).
 FT DOMAIN 3315 3403 COILED COIL (POTENTIAL).
 FT DOMAIN 3643 3666 COILED COIL (POTENTIAL).
 FT NP-BIND 1890 1897 ATP (POTENTIAL).
 FT NP-BIND 2169 2176 ATP (POTENTIAL).
 FT NP-BIND 2174 2181 ATP (POTENTIAL).
 FT NP-BIND 2520 2527 ATP (POTENTIAL).
 SQ SEQUENCE 4196 AA: 484 406 MW: 86104637.0184 P-00 QRC64;
 Query Match 2.7%; Score 109; DB 1; Length 4196,
 Best local Similarity 19.6%; Pred. No. 29;
 Matches 165; Conservative 123; Mismatches 286; Indels 268; Gaps 46,
 QY 85 KVNIAVERNFIPLGSLPLAEPFRNI-PL---LGRPPPTLQOITENLIKKGTHILLSAT 140
 Db QISKTALEENFW------QFYRSLRLNQILSRPVLVLD---ILAFGRFHTAS 323
 QY 141 LGGRHSITLIPVDK-RKLSKRAKGSUSFTTNSSSVTLTTHQIAASYPTIDKDS----- 190

DB 324 ENSEINIOCFVQKCKDALKKEISLIDFLSSSTLESILQLSAALIVSTFSKKWNTQYF 382
 QY 191 TLPPRLHHIQ-TASTAIKVTEPTPTPLGCSNYENLD-----SVSSVL----- 240
 DB 383 ETRVILDFINF-TEDILKILSRILPALGALISNVDFSHRTAVSSDILSLQYRIKIDFLRI 442
 QY 231VQSPENKIQGGIQLIPYIQFRRVCAALSYIACNSE--G 269
 DB 443 SLSKEEAGSYVGLKNSIKVIRAFENKRLK-----YIQS-FHEKHOQLIGALSEVYG 491
 QY 270 PPIKENDQWCHGCPKPFQNCFSM--DIQAMENMLRIT---ETWKA-----YNSDFE 318
 DB 492 LTHITELFELHUNKKEHVFNLITVFKDQSL--NVLDISLKGVNANSLTSSVNCMTV 549
 QY 319 ESEDFKLEKRLPMKYFLNTSTIMHWTMSNCFPRY--EOLNSMKO-LFLKAKIVHK 375
 DB 550 LHDVIAQLKSL-IOY-----SKTSSOMFTLMKQFPRFVRVTSISDCLHIAVNRIRK 605
 QY 376 LFSISKRGHK-----QPLISLPKORISIVWIKIQSLFY---CNENGLIGS----- 418
 DB 606 LDKLTRFEDVSDTELAMNELRLNPASSAILIWTOLKKLHEYTKNIIIEGEDWNN 665
 QY 419FSHETHSCOTCPNDQV 433
 DB 666 FIDGFEKVECTIQKRTDTNIIETNWINVSSPNINFEFESKIEFYITQSSAKSPRLS 725
 QY 434 VCTAFPLPTVGTASACLTCA-----PLNTRISQINIGYMLSGDLCKPEVAESTHYIG 487
 DB 726 VSDIDHVSF--CKRIPTLAHGYNIPQSLMEHASCIPQIL-AMCLIDSVQSPND VS 781
 QY 488 FETDQQLKMYLLQNTDRIEVH---AIFIS-NDMRINSWFDP----- 527
 DB 782 FEISKIE-EERFELQEYELAVRQHIVTGLFISWDFVGNLSIPEKCAIGKRNFLKNHP 840
 QY 528 -SWRKRMLITLKSNNKYSSIVHMLIGI-----SLQIC-----LTKNSTLEPVL- 569
 DB 841 NVENYVQFSLSITSLIMNK RNATSHYMQIQEGIPQADICEYSGDIFLIDRKLQQLID 899
 QY 570 AVYNPFGGSHSEWFMVPVNSFEDWERTKLDLPLOCYNNWTLILGNKWKTEFFETHYL 629
 DB 900 LIYVNGY-----SNLPPFVRA---LINRPODLIISKCKKFLSPFKIT----- 938
 QY 630 ASRIKSNCFNGNE-----SIYEPLEFIDFSRNIQYMAINNIOVGYSMHFDPAIR-- 681
 DB 939 ---ILTSGNENNOLKSKFSSDMYEKLK-----GFIKPTNLTIQNIIIEFDPVVEKK 987
 QY 682 ----DLILQ IDYIV TQGSQSAQLQILHGRWVVKLSIPGGRRLHFSCLL 728
 DB 988 EDSIVLLDMCIQSVDVNIPLISIKITTAQRNCHLIGFFIV---INKLE---SLHIGIFRSL 1041
 QY 729 RH 730
 DB 1042 FH 1043

Search completed: May 19, 2003, 10:32:02
 Job time : 40 secs

before version 1.1.1. 4578
Copyright (c) 1993-2003 by SeqAn Ltd.

SeqAn: using SeqAn API

2003.10.27.18: Search time 97 seconds
(without alignments)
(with 1.1% Million cell updates/sec)

432 129-35

SeqAn: PSLMA:WFW SeqAn: PNTMD:ITLCS 766

SeqAn: Gapext 0.5

SeqAn: 206047115 residues

SeqAn: chosen parameters 671580

SeqAn: 0.0000

SeqAn: 0.0000

SeqAn: 45 summaries

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

SeqAn: 45

ALIGNMENTS

RESULT 1

ID	Q95560	PRELIMINARY:	PRT:	781 AA
AC	Q95560			
DT	01-MAY-1999 (TrEMBLrel. 10, Created)			
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)			
DT	01-JUN-2001 (TrEMBLrel. 17, last annotation update)			
DE	Hypothetical 88.7 kDa protein (Unknown) (Fragment)			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo			
OX	NCBI_TaxID:9606			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Rhodes S.			
RL	Submitted (JAN-1979) to the EMBL/Genbank/DDBJ databases.			
RN	[2]			
RP	SEQUENCE OF 576-781 FROM N.A.			
RX	MEDLINE:96207227; PubMed-8519474;			
RA	Andersson B., Wentland M.A., Ricciarelli J.Y., Liu W., Gibbs R.A.			
RT	"A 'double adapter' method for improved shotgun library			
RT	construction."			
RL	Anal. Biochem. 236:107-113(1996).			
RN	[3]			
RP	SEQUENCE OF 576-781 FROM N.A.			
RX	MEDLINE:97264341; PubMed-9110174;			
PA	Yu W., Andersson B., Wentland M.A., Ricciarelli J.Y., Liu W.,			
PA	Ricciarelli J.Y., Wentland M.A., Lennon G., Gibbs R.A.			
RT	"Large scale concatenation cDNA sequence library			
RT	Genome Res. 7:353-358(1997).			
RN	[4]			
RP	SEQUENCE OF 576-781 FROM N.A.			
RA	Mei G., Yu W., Gibbs R.A.			
RT	Submitted (FEB 1997) to the EMBL/Genbank/DDBJ databases.			
DR	EMBL: A1035789; CAA22893.1			
DR	EMBL: A131933; AAU20055.1			
DR	InterPro: IPR000561; FGF-like			
DR	InterPro: IPR001179; FKBP_PPase			
DR	InterPro: IPR002384; GLA_bone			

number of results produced by chance to have a
in or equal to the score of the result being printed,
analysis of the total score distribution.

SUMMARY

SeqAn: ID	Description
Q95560	homo sapien
Q9C0B6	homo sapien
Q60477	homo sapien
Q920P3	mus musculus
Q925T8	rattus norv
Q9X10	mus musculus
Q961G0	homo sapien
Q95726	homo sapien
Q8W22	homo sapien
Q8W56	homo sapien
Q920P4	mus musculus
Q9J37	human herpe
Q9J35	arabidopsis
Q9UQ3	sus scrofa
Q9W33	human herpe
Q9XZ9	leishmania

DR InterPro: IPR001862; MAC_perforin.
 DR PRINTS: PR00002; GLABONE.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00457; MACPF; 1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 781 AA; DEE3EBB3A0RB599 CRC64;

Query Match 72.6%, Score 2962, DB 4, Length 791,
 Best Local Similarity 70.6%, Pred No Gaps 242,
 Matches 543, Conservative 99, Mismatches 109, Indels 18, Gaps 4;

QY 16 ALWEH---IALSLRHWLAVAA-----VSDQAT-----SPFDWLLSDKGPSPHRSQETDF 63
 DB 13 AVAPWATALLALGLPCWVLAVSATAAAVVPQHASVAGQHPDLDWLLTUNGPPHRAQYADF 72
 QY 64 VLRSGKSTSTPKYIPERIPKVNNAIAVERPNPIGSDPIAPERPNPIIGPPPIQOI 123
 DB 73 MERYQGFITPIYPIYEPAPKVNNAIAEPKDPESLPIAPERPNPIIGPPPIQOV 142
 QY 124 TENLIKKGTHPLSATLGGESLTIPIVDKRLKSRAGSDSTT-----NSSSVILETL 177
 DB 133 TENLIKKGTHPLSATLGGESLTIPIVDKRLKSRAGSDSTT-----NSSSVILETL 192
 QY 178 HQLAASYFIDRSTLRRLHHQIATATIKVTETPTSPUP'SNYINLDSVSVLVQSPENK 247
 DB 193 HQLAASYFIDRSTLRRLHHQIATATIKVTETPTSPUP'SNYINLDSVSVLVQSPENK 252
 QY 238 IQGLQGLVLLPVLQYQKPVQALSYIACNSGSPHCKPNCWCHGCKPFCPCNSMDIQ 297
 DB 253 VQLLGLQVLLPVLQYQKPVQALSYIACNSGSPHCKPNCWCHGCKPFCPCNSMDIQ 312
 QY 298 AMEENLRITETWKAYSNDFEEDKFLMKRLPMNYFLNTSTIMHLWTMDSNFORRYEQ 357
 DB 313 AMEDSLIQDSWATHNQPFEESEFOALIKRIPDGRFLNSTAISQFWAMVTSLOHRYQ 372
 QY 358 LENSMLQFLKAGKIVHKLFSLSKRCHKOPILISLPRTSTYWLTRIQSLFYCNENGLIG 417
 DB 373 LGAGLKVLFKFKTHRTLPRLFNFKPCHPQPPFLPKPEPSLYWNKRALQSLLYNSESTPG 432
 QY 418 SPSEETHSCPTCNQDVCTATPCTVGVGASACTACAPNTRCTCTNTIMHLWTMDSNFORRYEQ 477
 DB 433 TFEQSHSCPTCYDQSSCGGPIPCALGPGACACAPNTRCTCTNTIMHLWTMDSNFORRYEQ 492
 QY 478 VAISTDHYIGFTDIDQEMKYLQKTRTEVHAIFTISNDMLNSWPDYPSWRRKMLLTL 537
 DB 493 VAESLENFLGLETDLQDLKYLQKTRTEVHAIFTISNDMLNSWPDYPSWRRKMLLTL 552
 QY 538 KSNKYKSSLVHMTLGLSLQICLTKNSTLPIVAVVNPFGGSHSESWFMPVNSFPDWE 597
 DB 553 KSNKYKPGVLVHVMIALSLQICLTKNSTLPIVAVVNPFGGSHSESWFMPVNSFPDWE 612
 QY 598 RTKLDELPLQYQNTITLGNKWKIFFEIVHLYLRSRIKSNIPGNSESTIYYPELEFDPSRN 657
 DB 613 RTNVDAQAQCNWTITLGNKWKIFFEIVHLYLRSRIKSLDSSNETIYYPELEMTDESKN 672
 QY 658 LGYMKINNTQVENSMBHFLPEALRLLQLLYPYLQSSQSSALLQLLEIDPVRNKLSPPG 717
 DB 673 LGYMKINNTQVENSMBHFLPEALRLLQLLYPYLQSSQSSALLQLLEIDPVRNKLSPPG 732
 QY 718 QRRLDFSLRLRHRLKSTSEVVRISQALQAFNAKIPNTIMVITKIFS 766
 DB 733 KVRLLDFSLRLRHRLKSTSEVVRISQALQAFNAKIPNTIMVITKIFS 781

RESULT 2
 Q9C0B6 ID Q9C0B6 PPT, 791 AA
 AC Q9C0B6
 DT 01-JUN-2001 (Trembl et al. 17, Created)
 DT 01-JUN-2001 (Trembl et al. 17, Last sequence update)
 DT 01-JUN-2001 (Trembl et al. 18, Last annotation update)

DE KIAA1747 protein (Fragment).
 GN KIAA1747.
 OS Homo sapiens (Human).
 OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
 OC Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 PX MEDLINE=21082922; PubMed=11214970;
 RA Nagase T, Kikuno F, Hattori A, Kondo Y, Okumura K, Ohara O.
 RT Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.
 PL JNA RES. 7:347-355(2000).
 DR FMBL: AB051534; BAB21838.1; .
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001179; FKBP_PPIase.
 DR InverPro: IPR001862; MAC_perforin.
 DP SMART: SM00181; EGF; 1.
 DP SMART: SM00457; MACPF; 1.
 DP PROSITE: PS00453; FKBP_PPIASE_1; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 791 AA; 89887 MW; 21C5CD35AAF6A89 CRC64;

Query Match 72.6%, Score 2962, DB 4, Length 791,
 Best Local Similarity 70.6%, Pred No Gaps 242,
 Matches 543, Conservative 99, Mismatches 109, Indels 18, Gaps 4;

QY 16 ALWEH---IALSLRHWLAVAA-----VSDQAT-----SPFDWLLSDKGPSPHRSQETDF 63
 DB 23 AVAPWATALLALGLPCWVLAVSATAAAVVPQHASVAGQHPDLDWLLTUNGPPHRAQYADF 82
 QY 64 VLRSGKSTSTPKYIPERIPKVNNAIAVERPNPIGSDPIAPERPNPIIGPPPIQOI 123
 DB 83 MERYQGFITPIYPIYEPAPKVNNAIAEPKDPESLPIAPERPNPIIGPPPIQOV 142
 QY 124 TENLIKKGTHPLSATLGGESLTIPIVDKRLKSRAGSDSTT-----NSSSVILETL 177
 DB 143 TENLIKKGTHPLSATLGGESLTIPIVDKRLKSRAGSDSTT-----NSSSVILETL 202
 QY 178 HQLAASYFIDRSTLRRLHHQIATATIKVTETPTSPUP'SNYINLDSVSVLVQSPENK 247
 DB 203 HQLAASYFIDRSTLRRLHHQIATATIKVTETPTSPUP'SNYINLDSVSVLVQSPENK 262
 QY 238 IQGLQGLVLLPVLQYQKPVQALSYIACNSGSPHCKPNCWCHGCKPFCPCNSMDIQ 297
 DB 253 VQLLGLQVLLPVLQYQKPVQALSYIACNSGSPHCKPNCWCHGCKPFCPCNSMDIQ 322
 QY 298 AMEENLRITETWKAYSNDFEEDKFLMKRLPMNYFLNTSTIMHLWTMDSNFORRYEQ 357
 DB 323 AMEDSLIQDSWATHNQPFEESEFOALIKRIPDGRFLNSTAISQFWAMVTSLOHRYQ 382
 QY 358 LENSMLQFLKAGKIVHKLFSLSKRCHKOPILISLPRTSTYWLTRIQSLFYCNENGLIG 417
 DB 383 LGAGLKVLFKFKTHRTLPRLFNFKPCHPQPPFLPKPEPSLYWNKRALQSLLYNSESTPG 442
 QY 418 SPSEETHSCPTCNQDVCTATPCTVGVGASACTACAPNTRCTCTNTIMHLWTMDSNFORRYEQ 477
 DB 443 TFEQSHSCPTCYDQSSCGGPIPCALGPGACACAPNTRCTCTNTIMHLWTMDSNFORRYEQ 502
 QY 478 VAISTDHYIGFTDIDQEMKYLQKTRTEVHAIFTISNDMLNSWPDYPSWRRKMLLTL 537
 DB 503 VAESLENFLGLETDLQDLKYLQKTRTEVHAIFTISNDMLNSWPDYPSWRRKMLLTL 562
 QY 538 KSNKYKSSLVHMTLGLSLQICLTKNSTLPIVAVVNPFGGSHSESWFMPVNSFPDWE 597
 DB 563 KSNKYKPGVLVHVMIALSLQICLTKNSTLPIVAVVNPFGGSHSESWFMPVNSFPDWE 622
 QY 598 LGYMKINNTQVENSMBHFLPEALRLLQLLYPYLQSSQSSALLQLLEIDPVRNKLSPPG 717
 DB 623 RTNVDAQAQCNWTITLGNKWKIFFEIVHLYLRSRIKSLDSSNETIYYPELEMTDESKN 682
 QY 658 LGYMKINNTQVENSMBHFLPEALRLLQLLYPYLQSSQSSALLQLLEIDPVRNKLSPPG 717


```

Db 448 ATTAQKQKIQRTARKLFGLSVRCRHNPHQIPRETIQQWIAVQSLYCNENGFHGTFL 407
QY 421 EETHSTCPNDQVQVCTAFPLCTVGDASACLTGAPDNRTGCTGNTGMYLMSGLCKPEVAE 480
Db 408 ESQSCVWVHGSLTQKPLPCTVGDASACLTGAPDNRTGCTGNTGMYLMSGLCKPEVAE 467
QY 481 S--TQRYIGFETDL--QDLEMKYLLQKTDKRVHVAIFISNIMPLNSWFDPSWKKMLJLT 536
Db 468 SERSEVFSEFETDLQDLEKYLQKTDKRVHVAIFISNIMPLNSWFDPSWKKMLJLT 527
QY 537 LKSNKYKSLVIMLIGLSQICLTGKSTLEPVLAVVNPFGSGHSSESWMFVNENSEPDD 596
Db 528 LKSNKRNMDPIMHVIGMSKRIQUMRNSSDIMPFVYVNPFGSGHSSESWMFVNENSEPDD 587
QY 597 ERTKLDLPQCYNMILLTANKWKTFEFTVHIVLESPKISNIPNINFEISYIYPLEFLIPSR 656
Db 588 EKIRLQ-NSQCYNMILLTANKWKTFEFTVHIVLESPKISNIPNINFEISYIYPLEFLIPSR 644
QY 657 NIAGYMKINNIQVFGYSMBFDEPAIRQLTIGDHYPTQCSQ----DSALLQLLEIRKVN 712
Db 645 RQFYIKISDVQVFGYSMBFDEPAIRQLTIGDHYPTQCSQ----DSALLQLLEIRKVN 704
QY 713 LSP---PGQPRIDLPSCILPHLKLKSTSEVPLQALQAFNAKLPTNDYDTTKLC 765
Db 705 LAPPVAPGKQDLPSCKLHKRLKINSEIRVNHALDIYNTEILKQSDQMTAKIC 760

RESULT 5
Q925T8
ID Q925T8 PRELIMINARY: PRT: 760 AA.
AC Q925T8
DT 01-DEC-2001 (TRENBLrel, 19, Created)
DT 01-DEC-2001 (TRENBLrel, 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel, 19, Last annotation update)
DE BMP/retinoic acid-inducible neural-specific protein.
GN BKNP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Patius
OX NCBI_TaxID=10116;
RN [1]
RP SOURCE FROM N.A.
RA Asari M.; Nakatani T.; Kawano H.; Kobayashi M.; Abe A.; Araki N.;
RA "Identification and characterization of RPNP, a BMP/retinoic acid
RT inducible neural-specific protein."
RL Submitted (NOV 2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051356; BAB55642.1;
SU SHQUNCK 760 AA; 88654 MW; 2243964946C265D3 CRC64;

Query Match 52.4%; Score 2136.5; DB 11; Length 760;
Best Local Similarity 51.7%; Pred. No. 6.8e-172;
Matches 402; Conservative 133; Mismatches 213; Indels 29; Gaps 10;

QY 1 MWKSRAGALFLSMLALWEIALLSLHCWVLAVAASDQHATSPDMLISDKGPHRSQF 59
Db 1 MNRRF---VELLYFLVWGRISVQPSH----QEPAGTQHVSKPEEDWLDISDRGPHHSRS 53
QY 60 YTDVDFSPGSGFTPYKTYPEMSPKVNNAVERPFLGSLPLAPEDFPNIRPLIGRPPT 119
Db 54 YLSFVERHHQGFTRYKYREAPKVRNTATERRDLVRHVPVLMPEPQRSIRILGRPT 113
QY 120 LQGITENLKKYGTHTPLLSATIGTRESITFVKKKLSKKAEGSDSTINSSVILETLHQ 179
Db 114 TQGFIDITIKKYGTHTPLLSATIGTRESITFVKKKLSKKAEGSDSTINSSVILETLHQ 166
QY 180 LAASYFIDRINSLPLRHHQIATAIKVTEPTGPGCSNYDNLSVSSVLQSPENKIQ 239
Db 167 LASSYFVDRGTMRLRHEIQISTGAKVTFTPTGLGCSNDNLSVSSVLQSTESKLI 226
QY 240 LQGITVLLPDYQERFQVAAVLSIACNSEGEFLKKNKDWVHGGKPEFCNCPMSDIQAM 299
Db 227 LQGITVLLPDYQERFQVAAVLSIACNSEGEFLKKNKDWVHGGKPEFCNCPMSDIQAM 286

```

```

QY 360 PLNLJLRLTAAAYNSDRESDPKLPKRLMAYELNLSLIMHWMUSANEQRYTQDE 359
Db 287 EFTLMAKAWTAYTQDLENSEDEFSKPKLPSPNIPTTFTATHTHWNWPLQAPVPLIQ 446
QY 360 NSMKOIFLKAOKLVHKLPSLSKRCCHKOPLISLPORSTSYWLTRIOSPLYNCPNCLGSP 419
Db 347 SAIEAGKOKIQRIAKLPLGLSVRCRHNPHQIPRETIQQWIAVQSLYCNENGFHGTFL 406
QY 420 SEETHSTCPNDQVQVCTAFPLCTVGDASACLTGAPDNRTGCTGNTGMYLMSGLCKPEVA 479
Db 467 ESQSCVWVHGSLTQKPLPCTVGDASACLTGAPDNRTGCTGNTGMYLMSGLCKPEVA 466
QY 480 PS--TQRYIGFETDL--QDLEMKYLLQKTDKRVHVAIFISNIMPLNSWFDPSWKKMLJLT 535
Db 467 SERSEVFSEFETDLQDLEKYLQKTDKRVHVAIFISNIMPLNSWFDPSWKKMLJLT 526
QY 536 NIAGYMKINNIQVFGYSMBFDEPAIRQLTIGDHYPTQCSQ----DSALLQLLEIRKVN 711
Db 644 RQFYIKISDVQVFGYSMBFDEPAIRQLTIGDHYPTQCSQ----DSALLQLLEIRKVN 704
QY 712 LSP---PGQPRIDLPSCILPHLKLKSTSEVPLQALQAFNAKLPTNDYDTTKLC 765
Db 704 LAPPVAPGKQDLPSCKLHKRLKINSEIRVNHALDIYNTEILKQSDQMTAKIC 760

RESULT 6
Q90X10
ID Q90X10 PRELIMINARY: PRT: 761 AA.
AC Q90X10
DT 01-MAY-2000 (TRENBLrel, 13, Created)
DT 01-MAY-2000 (TRENBLrel, 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel, 19, Last annotation update)
DE DHCCR1.
GN DHCCR1 OR DHCCR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SOURCE FROM N.A.
RA STRAIN=C57; TISSUE=BRAIN;
RA Ochal L.K.; Sowden M.P.; Messing E.M.; Wheelless L.L.; Reeder J.E.;
RT "Mouse Dhccr1 cDNA."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF202896; AAF17579.1;
DR MGI; MGI:1928478; Dhccr1.
DR Interpro; IPR001862; MAC_perforin.
DR SMART; SM00457; MACPF; 1.
SQ SEQTENFE 761 AA; 88750 MW; F2792TA84FBG8C9B CRC64;

Query Match 52.1%; Score 2128; DB 11; Length 761;
Best Local Similarity 51.7%; Pred. No. 3.6e-171;
Matches 402; Conservative 131; Mismatches 216; Indels 28; Gaps 10;

QY 1 MTWSPAPAFLEFLSMLALWEIALLSLHCWVLAVAASDQHATSPDMLISDKGPHRSQF 60
Db 1 MNRRF---VELLYFLVWGRISVQPSH----QEPAGTQHVSKPEEDWLDISDRGPHHSRS 54
QY 61 YTDVDFSPGSGFTPYKTYPEMSPKVNNAVERPFLGSLPLAPEDFPNIRPLIGRPPT 120
Db 55 LSEFVERHHQGFTRYKYREAPKVRNTATERRDLVRHVPVLMPEPQRSIRILGRPT 114
QY 121 LQGITENLKKYGTHTPLLSATIGTRESITFVKKKLSKKAEGSDSTINSSVILETLHQ 180

```


QY 504 TDRFIEVHAIFISNOMPLNSWFDPSWPKRMILLTKSNK 541
 DB 181 TDRFIEVHAIFISNOMPLNSWFDPSWPKRMILLTKSNK 218

RESULT 9

Q8WU22 Q8WU22 PRELIMINARY: PRT: 320 AA.
 AC Q8WU22:
 DT 01-MAR-2002 (TREMBLrel, 20, Created)
 DT 01-MAR-2002 (TREMBLrel, 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
 DE Similar to deleted in bladder cancer chromosome region candidate 1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RA Strausberg R.;
 RL Submitted (JAN 2002) to +the EMBL/GenBank/DBJ databases.
 DR EMBL: BC021560; AAI21560.1;
 DR InterPro: IPR001862; MAC_perforin.
 DR SMART: SM00457; MACPF; 1.
 SQ SEQUENCE 320 AA: 37168 MW: 21644CH38764P5G3 CRG64.

Query Match 23.8%; Score 970.5; DB 4; Length 326.
 Best Local Similarity 56.4%; Pred. No. 9,6e-74;
 Matches 181; Conservative 53; Mismatches 72; Indels 15; Gaps 4;

QY 1 MIWRSRAGAEPLFSIMALWEWIALS-LHCWVLAAVAASVDHATSPFDWLLSDKGFPHRSQS 59
 DB 1 MNWRF---VELLYFLFWGRTISQVPSH---ORPACTDQHSKKEFDWLLSDKGFPHRSKS 53
 QY 60 YTFDVSROGCTFTRKYIPEFAPKVNNAIYVPPNFISPIPIAPEFPNPPIIGRPPT 119
 DB 54 YLSEVHRHOGCTFTRKYIPEFAPKVNNAIYVPPNFISPIPIAPEFPNPPIIGRPPT 113
 QY 120 LQQTENILKKYTHLLSATIGGHSFIVKPKKSPAGSDSTINSSSVTLIELHQ 179
 DB 114 TQGFDTILKKYTHLLSATIGGHEALTYMDKSRLE-----DRKSNATQSVEALHQ 166
 QY 180 LAASYFDINRSTPLRIHHTQIASTAIKVTFTFTRGAGSNYINLDSVSVLVQSPENKIQ 239
 DB 167 IASSYFDVDRGIMKRLHRLQISTGAIKVTFTFTRGAGSNYINLDSVSVLVQSPENKIQ 226
 QY 240 LQQTENILKKYTHLLSATIGGHSFIVKPKKSPAGSDSTINSSSVTLIELHQ 299
 DB 227 LQQTENILKKYTHLLSATIGGHEALTYMDKSRLE-----DRKSNATQSVEALHQ 286
 QY 300 PENILRITETWKAYNDSDFES 320
 DB 287 EYTLANNAKSWAEAYKDLNS 307

RESULT 10

Q8WX56 Q8WX56 PRELIMINARY: PRT: 228 AA.
 AC Q8WX56:
 DT 01-MAR-2002 (TREMBLrel, 20, Created)
 DT 01-MAR-2002 (TREMBLrel, 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
 DE RA2804.1 (deleted in bladder cancer chromosome region candidate 1 (B3089A)) (Fragment).
 GN DBCCR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Babbage A.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AII38894; CAD13302.1;
 DR InterPro: IPR001862; MAC_perforin.
 DR SMART: SM00457; MACPF; 1.
 FT NON_TER 228 228
 SQ SEQUENCE 228 AA: 25585 MW: 1878869E491774I4 CRG64.

Query Match 17.7%; Score 720.5; DB 4; Length 228.
 Best Local Similarity 57.9%; Pred. No. 8,5e-53;
 Matches 140; Conservative 40; Mismatches 47; Indels 15; Gaps 4;

QY 1 MIWRSRAGAEPLFSIMALWEWIALS-LHCWVLAAVAASVDHATSPFDWLLSDKGFPHRSQS 59
 DB 1 MNWRF---VELLYFLFWGRTISQVPSH---ORPACTDQHSKKEFDWLLSDKGFPHRSKS 53
 QY 60 YTFDVSROGCTFTRKYIPEFAPKVNNAIYVPPNFISPIPIAPEFPNPPIIGRPPT 119
 DB 54 YLSEVHRHOGCTFTRKYIPEFAPKVNNAIYVPPNFISPIPIAPEFPNPPIIGRPPT 113
 QY 120 LQQTENILKKYTHLLSATIGGHSFIVKPKKSPAGSDSTINSSSVTLIELHQ 179
 DB 114 TQGFDTILKKYTHLLSATIGGHEALTYMDKSRLE-----DRKSNATQSVEALHQ 166
 QY 180 LAASYFDINRSTPLRIHHTQIASTAIKVTFTFTRGAGSNYINLDSVSVLVQSPENKIQ 239
 DB 167 IASSYFDVDRGIMKRLHRLQISTGAIKVTFTFTRGAGSNYINLDSVSVLVQSPENKIQ 226
 QY 240 LQ 241
 DB 227 LQ 228

RESULT 11

Q920P4 Q920P4 PRELIMINARY: PRT: 72 AA.
 AC Q920P4:
 DT 01-DEC-2001 (TREMBLrel, 19, Created)
 DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
 DE BMP/retinoic acid inducible neural specific protein (BRINP) (Fragment).
 GN BRINP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090,
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EMBRYONIC STEM CELLS;
 RA Matsumoto K., Ariga H.;
 RA Matsumoto K., Nakatani T., Kawano H., Arai N., Abe A., Kobayashi M.;
 RT "Identification of BMP/RA-inducible neural specific protein."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB060587; BAB70660.1;
 FT NON_TER 72 72
 SQ SEQUENCE 72 AA: 8838 MW: 62CDP1B03256AR2D CRG64.

Query Match 4.3%; Score 175; DB 11; Length 72.
 Best Local Similarity 46.2%; Pred. No. 2,4e-07;
 Matches 36; Conservative 11; Mismatches 25; Indels 4; Gaps 2;

QY 1 MIWRSRAGAEPLFSIMALWEWIALS-LHCWVLAAVAASVDHATSPFDWLLSDKGFPHRSQS 60
 DB 1 MNWRF---VELLYFLFWGRTISQVPSH---ORPACTDQHSKKEFDWLLSDKGFPHRSKS 54
 QY 61 TDFVDRSGSFTRYKIY 78
 DB 55 LSPVHRHOGCTFTRKYIY 72

RESULT 12

Q9QJ37

MINARY: PRI: 200 AA
Seqrel: 13, Created)
Seqrel: 13, Last sequence update)
Seqrel: 20, Last annotation update)
56.
Fused, no RNA stage: retroviridae;
: Roseoviridae.
PubMed:10482533;
Smith T.R., Staley F.E., Jawhurst S., Inoue N.,
8 kb genome sequence: coding content and comparison
Zelus SA.
605221000)
AD4945.13
5: MYB DNA binding
MYB11 UNKN/WN
A: 240284 MW: 490474 JAFKUF8 CR064;
y 228; Score 129; 16 12; Length 2077;
y 21.0%; Pred. No. 125;
yative 109; Mismatches 248; Indels 242; Gaps 44;
26STRVYKIVREFGKRVNNA --VERKFLGSLPLAPEFFNNI 111
NKET--KIVD-----NYNLAAGHMTVLNGLFAYL---VNLVNI 633
LHQULNI-----IKKVF HF --LSAALGGHSLTFVVKKL 156
ELLGFVKIYNTGTFRTSHNINNLSEFISIRKMEYDFYNSQRP 693
--SESTINSSVILELHQALAAVEIDRSILPLHLHLIASIAIKV 207
SLAIPSKNKLIVNEM *SQIFELGKMLSLTENKO 742
LGSNYNILSVSS-----VLVSPE-----NKIQ 240
BLTFYFLHLISAVN ENIVVNLHLHFLKPPDIAAILHDKIQ- 801
RRPV QAAISYIACNSGDEFLKNIW*WCHCKPKPCNCHSMUQAM 299
GIVFEQTESIVL ELV -----DLFPN-----STEKIAL 841
AYNSFEETEFLEFMYKELIMAYVNIIMILWTMDSNFPVPVQ 357
LAL 877
LQVVERFESYKPKHKKLQICLQALPTSTWITPTISFYNNELIG 417
KFAFLIMHEKELYSQFHLAEN --WLPQ----- 416
NGVVVAFATFIVDAATGAGAN--KSNVNGYMSGCKPKH 477
RE--LTAFL-----ATAPKHALQ--TKPE 947
TDLAQEMKYIQLTGGFVVAATEIN LPLNSWEDPSWPKMIL 535
--MEQMKOTAGNKKHLLTENLEKRLND-----LIL 987
-----SLVIMILGLSLQCLTKNSTLEPV 568
VILNLETELQALUNL LHEHALFVTKHETTLTKTISGLEKI 1047
HSESWFMVNSFDMFKLHQLQYNNILITLGNKW-KIFFEIVH 626
KEE.PYTH-----LSTFKPKKILPT-- 1082

QY 627 IYLSPIKSPNGNESIYEPLEFIDPSNLAQYMKINNIOVRYSMIFDPAIPILIQ 686
Db 1083 ----EDVKTEIDQMSKSTF--LTKIAISIN-KYTRISH-SVYQJLNLYEERITELKKE 1134
QY 687 LHWVYI26SDSALIQ---LPIRD-KVAKISDQGGKRRIDFSCILLRHHL----- 742
Db 1135 TNKIKKESKEYAVAEKKILLSSQDAFNK-----IYLVNTHTLKRIKNTQFRE 1184
QY 733 ----KLSISIV---VRIQALQAFNAKLIQNTIMDYD-11K: 764
Db 1185 TAFAKALIVEVNNKENQLOHLLNHNFAHLKAKMKQNH11K1.225
RESULT 13
Q9LJTS PRELIMINARY: PRI: 529 AA.
AC Q9LJTS;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-77N-2002 (TrEMBLrel. 21, Last annotation update)
DE GBAAAD17333.1
OS Arabidopsis thaliana (Mouse-ear cross);
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu F., Tabata S.;
PL Submitted (Sep-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
PX MEDLINE=20363099; PubMed=10907853;
PA Nakamura Y.;
RT *Structural analysis of Arabidopsis thaliana chromosome 4, II.
PT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
KT IAC and HAC clones.
PL UNA POS 7:217-221(2000).
DR EMBL: AP000411; HAK02137.1;
DR InterPro: IP001849; Gram_pos_anchor.
DR InterPro: IP005162; Retrotrans_qag.
DR InterPro: IP004364; tRNA-swt.2.
DR Pfam: PF03732; Retrotrans_qag.1.
DR Pfam: PF00152; tRNA_swt.2.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKN/WN 1.
SQ SEQUENCE 529 AA: 59477 MW: 884088BBAER9 CR064;
Query Match 3.1%; Score 127.5; Db 10 Length 529;
Best local Similarity 21.0%; Pred. No. 0.046;
Matches 107, Conservative 67, Mismatches 248, Indels 242, Gaps 46;
QY 161 TLPLHLHLHLIASIAIKVTFETPTFLPSNLYNLSVSVLVSPENKFLQCLQVLLIY 250
Db 73 TLPLHLHLHLIASIAIKVTFETPTFLPSNLYNLSVSVLVSPENKFLQCLQVLLIY 120
QY 251 TLPLHLHLHLIASIAIKVTFETPTFLPSNLYNLSVSVLVSPENKFLQCLQVLLIY 419
Db 121 -----LVNDSHGEEQVED-----RVSEPS-----DENLRCLSNF 151
QY 311 KAYNS-----DFFSFFFLFELFELPMNLYNLSIMILWIMDSNFPVRYE--QLENSMK 363
Db 152 ESYCCENNIPEHSWKPKKMMAR-----EKI-----IMNVNHQPHYSQIQHGSVR 199
QY 364 QLFKA*IVHKLPSLSKPK*HKUPLISLPKPTSTWITPTISFYNNELIG 418
Db 200 EYPERFEAL-----G--LSLVLPQCHLEETFLQKPLQLDAVPRFNSIVGM 247
QY 419 FS-----EHRSCIQPMQVVCIAFLPICVGLASA-----GICAPGNRTPC 460
Db 248 MNLAQWLEANE-----NEN-----GQVETRAVSVYKA--ELLMNTMTVQNGEVT 295

```

QY 461 GTCNTGCM:SGGCGKCFEVALSTDRYIGFEDLQLEEMKYLQKIDRRKEVHALFTSNMVK 520
DB 296 STC-----GLMK-----NKLEKFF-----QSSGHHVSVRIDSG 324
QY 521 LNSWFDPSRKRMLTLKSNKYKSSVIMILGLSLQJCLTKNSTLEPVLAVVYNPFGGSH 580
DB 325 AUNNMP-----KDIATHPKDKETNIVSVLGHGHLF-----K 359
QY 581 SESWEPVNVENSEPWPETKLDLPLOQYNWTLTUSNKKWKTFFE---TVHLYLR-----SR 632
DB 360 SKEKCMIGTSTI-NWERN-----TMSNH-----GNKVVITGEENDVHLHPYTKENLKLS 409
QY 633 IKSNGPNGNESVYEPLEFIDPSRNIGYMKINNIQVFG 670
DB 410 IRYKMSGNEQ-----BELNKTAREGPAEENSWKQEG 441

RESULT 14
QYUQ3 PRELIMINARY: PRT: 843 AA.
AC Q9TUQ3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Complement component C7 precursor.
GN C7.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A., SEQUENCE OF 640-651, AND TISSUE SPECIFICITY.
RP TISSUE=KIDNEY;
RC MEDLINE=20341712; PubMed=10878384;
RA Agha A., Montalto M.C., Kiesecker C.L., Morrissey M., Grover M.,
RA Whoolery K.L., Rother R.P., Stahl G.L.;
RT Isolation, characterization, and cloning of porcine complement
RT component C7.
RL J. Immunol. 165:1059-1065(2000).
CC 1- FUNCTION: C7 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX. C7
CC BINDS TO C5B FORMIN; THE C5B-7 COMPLEX, WHERE IT SERVES AS A
CC MEMBRANE ANCHOR.
CC 1- SUBUNIT: MONOMER OR DIMER; AS A C5B-7 COMPLEX IT CAN ALSO FORM
CC MULTIMERIC ROSETTES (BY SIMILARITY).
CC 1- TISSUE SPECIFICITY: BONE MARROW, HEART, INTESTINE, LUNG, SPLEEN,
CC KIDNEY, LIVER AND THYMUS.
CC 1- PTM: C7 HAS 28 DISULFIDE BRIDGES (BY SIMILARITY).
CC 1- SIMILARITY: TO COMPLEMENT FACTORS C6, C8, C9, AND TO PERFORIN.
CC 1- SIMILARITY: CONTAINS 2 SUSHI (SCR) REPEATS.
CC 1- SIMILARITY: CONTAINS 2 TYPE-1 TSP REPEATS.
CC 1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC 1- SIMILARITY: CONTAINS 1 LDL-PEPTIDE CLASS A DOMAIN.
EMRL: AF162274, AAD47418 1.
DR HSP90: Q07954; ICR8.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003884; FacI MAC.
DR InterPro: IPR002172; LDL_recept.A.
DR InterPro: IPR001862; MAC_perforin.
DR InterPro: IPR000436; SushI_SCP_Ccp.
DR InterPro: IPR000884; TSP1.
DR pfam: PF00057; ldl_recept_a; 1.
DR pfam: PF01823; MACPF; 1.
DR Pfam: PF00084; sushI; 2.
DR Pfam: PF00090; tsp_1; 2.
DR PRINTS: PR00764; COMPLEMENTC9.
DR SMART: SM00032; Ccp; 2.
DR SMART: SM00057; FIMAC; 2.
DR SMART: SM00192; LDLa; 1.
DR SMART: SM00457; MACPF; 1.
DR SMART: SM00209; TSP1; 2.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01209; LDLRA_1; 1.

```

```

DR PROSITE: PS00029; MAC_PERFORIN; 1.
DE PROSITE: PS00032; TSP1; 2.
KW Complement pathway; Complement alternate pathway: Glycoprotein;
KW Plasma; Membrane attack complex, Cytolysis, Signal, EGF like domain;
KW Repeat; Sushi; Transmembrane.
PT SIGNAL 1 22 HY SIMILARITY.
PT CHAIN 23 843 COMPLEMENT COMPONENT C7 (BY SIMILARITY).
PT REPEAT 24 83 TYPE-1 TSP 1.
PT DOMAIN 83 121 LDL-RECEPTOR CLASS A.
PT TRANSMEM 271 287 POTENTIAL.
PT TRANSMEM 292 311 POTENTIAL.
PT DOMAIN 451 487 EGF-LIKE.
PT REPEAT 497 545 TYPE-1 TSP 2.
PT DOMAIN 570 680 2 X SUSHI (SCR) REPEATS.
PT REPEAT 570 627 SUSHI 1.
PT REPEAT 630 689 SUSHI 2.
PT DOMAIN 695 770 COMPLEMENT CONTROL FACTOR 1 MODULE.
PT DOMAIN 771 843 COMPLEMENT CONTROL FACTOR 1 MODULE.
PT DISULFID 85 96 HY SIMILARITY.
PT DISULFID 91 109 BY SIMILARITY.
PT DISULFID 103 119 BY SIMILARITY.
PT DISULFID 337 353 BY SIMILARITY.
PT DISULFID 571 613 HY SIMILARITY.
PT DISULFID 599 626 HY SIMILARITY.
PT DISULFID 631 673 BY SIMILARITY.
PT DISULFID 659 688 BY SIMILARITY.
PT CARBOHYD 302 302 N-LINKED (GLCNAC...) (POTENTIAL).
PT CARBOHYD 754 754 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 843 AA; 93061 MW; 06703094C4P5CR4D CPO64;

Query Match 3.18; Score 127; DB 6; Length 843;
Best Local Similarity 20.58; Pred No 0.1;
Matches 100; Conservative 56; Mismatches 153; Indels 178; Gaps 23;

QY 55 HRSQYTD-----PVDRSRGCFSTRYKLYRFGWRKVNNI-----AVERNRITGSP--- 100
DB 210 HFTSTHTSSSGRVKIFSSSSSSSYAKYVHILKKSYQILVYVONTVRVAQFINNPEP 269
QY 101 LPLAPELLRPILOPPTLQOIT--ENLKKYGTHTLLSATLGGESLTITFYDKRKLK 158
DB 270 LQALSPFKWELSYL---PPLYDYSAYRRLIDYGTHTYLSQSGICGKYKVIIFYVDSKVAK 326
QY 159 RRGSDNTTNSSTVLEFTHOLAASYFIDRSTLPRHLHITQIASTAIK-----VITRTCP 214
DB 327 SDLGSEDKKKGASSHISFLFKSKHKCKAMEALKSASGTQ---SNVLRGVPEVRGGPRGF 384
QY 215 LQCSNYDNLDSVSSVIVGSPENKIQLQGLQVILGHNTLQERFVQAAALSYIACNSKGRPF 274
DB 385 VSGI.SYLELDN-----PDKRKQY----- 403
QY 275 ENDCWCHGCPKTECNCFSDIOAMEENLRLITETWKYNSDFEESDEPKLPMKRLPMNY 434
DB 404 ---SSWAGSVTDLP-----QVTKOKITPLYE-----LVKEVPC--- 433
QY 335 FLNTSTIMHLWTMDSNFGKRYEQLNSMKQLPILKAKILVHKLFSLSKRCHKQPLISLPRO 394
DB 434 -----ASVKRLDYK--RALEEYLDFTDSCHCQP----- 459
QY 395 RISTYWLIRIQSFILCYNHNGLLSGFSFETHSCDQ---PNDVVVCTAFILPC---LVCDAS 447
DB 460 -----C-ONG--GNASVEGTCQCHCKPN-----TFCVACRGGVIVGDHA 496
QY 448 -----ACLT-----CAPDNTRCGTONTGYMLSQG-LCKPEVAENSTDIHIGFTDLQDL 495
DB 497 GGIAGGWSWSSWGPGCAUGKKTRKCNKNNPPSGGKNSCIGETISER-----QCDEEDL 550
; QY 496 EMKYLILQ 502
DB 551 EHLRLLE 557

RESULT 15

```